The progressive in present-day spoken English

Real-time studies of its spread and functional diversification

Udo Rohe

NIHIN studies
New Ideas in Human Interaction

The progressive in present-day spoken English

Real-time studies of its spread and functional diversification

Inaugural-Dissertation
zur
Erlangung der Doktorwürde
der Philologischen Fakultät
der Albert-Ludwigs-Universität
Freiburg i. Br.

vorgelegt von

Udo Julius Rohe

aus Schwäbisch Hall

Sommersemester 2018

Erstgutachter: Prof. Dr. Dr. h.c. Christian Mair
Zweitgutachter: Prof. Dr. Dr. h.c. Bernd Kortmann
Drittgutachter: Prof. Dr. Dr. h.c. Peter Auer

Vorsitzender des Promotionsausschusses
der Gemeinsamen Kommission
der Philologischen und
der Philosophischen Fakultät: Prof. Dr. Joachim Grage

Datum der Disputation: 20.02.2019

Acknowledgements

This book – a revised version of my PhD thesis – was written between the years 2015 and 2018. At the beginning of the project, my father-in-law used the German phrase *Ein dickes Brett bohren* to describe the nature of a PhD. With hindsight, I could not think of a more fitting description. While the task certainly was a beneficial and fulfilling one, it also included aspects and periods that demanded a good deal of self-discipline and stamina.

I certainly would not have successfully completed this project without the help of a number of people. Above all, I would like to thank my supervisor, Christian Mair, for his constant and reliable support. His criticism was always constructive and productive. Further thanks go to all supervisors of the research training group DFG 1624 "Frequency effects in language". Especially to Bernd Kortmann, Peter Auer, Lars Konieczny and Stefan Pfänder, who were involved at different stages of the project. Equally important was the help of our group's coordinators Michael Schäfer, Nikolay Hakimov and Agnes Schneider.

Stephanie Horch, David Lorenz and Michael Percillier taught me how to work with *R* – thank you very much! Göran Köber did a great job in explaining different statistical techniques, and Isabel Schmier reliably and untiringly transcribed countless texts for my new corpus.

The summer school "Digital Humanities" 2015 in Leipzig and a research stay at the SFB 1102 "Information Density and Linguistic Encoding" in Saarbrücken also contributed to my methodological training (apart from being a lot of fun).

My fellow PhD students at the research training group – especially Sophia Wulfert, Stephanie Horch, Laura Terassa, Katja Schwemmer, Karolina Rudnicka and Carmela Pietropaolo – have made the PhD period a very sociable time and have become friends over the years.

I am also indebted to my proofreaders (Miriam Rohe, Jenny Baumann, Lisa Kirjalainen, Sophia Wulfert), who helped me a lot in the late stages of the project.

Furthermore, I would like to thank my parents, who have always supported my academic education.

Last but not least, I would like to mention my closest friend and companion: Miriam – my wife. Her positive personality and intelligence have been of invaluable support over the past years.

Contents

List of Figures

List of Tables

1. Introduction

Writing another book-length study on the English progressive (*be* + *-ing* participle) may appear as a somewhat daring endeavour. Given the amount of literature that has been published on the topic, it seems sensible to ask if anything of true importance remains to be said. Kranich (2010), Smitterberg (2005), and Núñez Pertejo (2004) studied different aspects of the progressive's historical development in great detail, while Mair and Hundt (1995) and Smith (2005) illuminated the construction's recent development in 20th-century English. Yet another approach was adopted by Römer (2005), who focused on the progressive's lexical-grammatical properties in spoken English. Finally, there is Rautionaho's (2014) recent work on progressive usage in World Englishes. Clearly, these and other studies make the progressive one of the best-researched constructions in the English language.

While this situation can be seen as limiting the potential scope of further research, it can also be regarded as offering many "loose ends" and open questions, some of which deserve closer attention. First of all, there is the issue of the progressive's current state of development. While it has been shown that the construction has rapidly increased in frequency in the second half of the 20th century, it is unclear whether the development is continuing in present-day English; and if yes, in which genres? In addition to that, the driving forces behind the frequency increase are still a matter of debate. On the one hand, linguists have argued for an expansion of the progressive's paradigm, especially with stative verbs (e.g. Levin 2013) and with supposedly innovative – non-aspectual – functions (e.g. König 1995a; Celle and Smith 2010; Pfaff et al. 2013). On the other hand, Leech et al. (2009) have put the focus on conventionalised contexts of use, most importantly the present progressive, arguing that it is exactly in these contexts that the progressive is spreading most dramatically. Based on this line of argument, the construction would be advancing under its own momentum, its most frequent uses being the (cognitively) most central ones, resulting in even more frequent usage (2009: 269f.). Since both explanations are based on empirical observation, it seems unlikely that we are dealing with a question of either/or. Therefore, the present work aims to quantify both the extent to which the conventionalised as well as the non-conventional/innovative contexts of use are developing. Invoking De Smet's (2016) concept of language change as an interplay of convention and innovation, it is proposed that both aspects of the progressive's recent development can meaningfully be reconciled with each other.

Finally, Römer's (2005) claim that the progressive is not a purely grammatical but a lexical-grammatical phenomenon raises a whole range of interesting questions – especially if a diachronic perspective is adopted. As Römer's work has shown, many recurrent progressive patterns develop pragmatic functions. For example, *I'm just wondering* frequently acts as a hedge, conveying a sense of politeness (2005: 126). By focusing on

selected patterns and studying their recent development in spoken English, the present work takes Römer's synchronic work one step further. This way, it will be shown how some of the progressive's most commonly used lexical-grammatical realisations have developed in speakers' everyday language over the past twenty years. In addition to the diachronic focus, close attention will be paid to the pragmatic side of meaning, accounting for the insight that repeated use in specific contexts may result in pragmatic strengthening and – ultimately – constructionalization[1] (cf. Bybee 2006 & 2007).

Thus, instead of discouraging further research, previous work on the progressive has resulted in at least three research questions that need to be addressed if one aims to provide a comprehensive account of the construction's present-day use and recent development:

1. Is the progressive's frequency development continuing in the 21st century? If yes, in which genres?

2. Which developments can be identified as the driving forces behind a potential increase – conventionalised aspectual uses, or non-conventional/innovative uses?

3. How have frequent lexical-grammatical progressive patterns developed in the course of the past twenty years? Do they consistently express routinised pragmatic meanings?

Focusing on spoken English, where changes are generally assumed to manifest themselves first, the present work warrants that the actual state of the development is captured. To do so, it was necessary to extend the existing range of spoken corpora by developing a new genre-specific corpus of spoken British English – the *Freiburg Corpus of Spoken English* (FCSE), which acts as a supplement to the existing *Diachronic Corpus of Spoken English* (DCPSE) (Aarts and Wallis 2006). Additionally, it was imperative to make use of the much larger *British National Corpus* (BNC) and its recent extension, the Spoken BNC2014 (Love et al. 2017) to obtain enough data for the analysis of lexically specific patterns. This deliberate combination of different corpus resources provides a sound empirical basis for the study of the progressive in contemporary spoken English.

Methodologically, this work resorts to a number of different statistical techniques, ranging from basic significance testing, over corpus-linguistic tools such as Collostructional Analyis (Stefanowitsch and Gries 2003), to logistic regression modelling (Gries 2013b) and visualisation by means of mosaic plots (Friendly and Meyer 2016). In doing so, it is not only possible to focus on the development of individual determinants of progressive use but to analyse (the interplay of) several factors at the same time.

The conceptual framework against which the results are interpreted is a usage-based (cf. Barlow and Kemmer 2000; Langacker 2000), constructionist (cf. Goldberg 2006)

[1] Even though the present work generally uses British English spelling variants, conventionalised or frequently used linguistic terms such as *grammaticalization*, *colloquialization* or *constructionalization* are spelled with a *z*. This also applies to the terms *schematization* and *categorization*, which both refer to specific usage-based concepts. However, it does not apply to terms that refer to less fixed concepts; e.g. *routinisation*.

one. Assuming that grammar is shaped by usage and based on domain-general cognitive processes (cf. Bybee 2010), speakers' knowledge of the progressive construction is conceived of as a complex exemplar-based network, whose nodes vary with regard to their degree of cognitive entrenchment and routinisation in the speech community. One important variable in this respect is frequency of use: While normalised token frequencies are assumed to be an indicator for the strength of mental representation, relative frequencies can help to determine which realisations of a variable are the most/least typical ones. Type frequencies, on the other hand, are often regarded as a proxy of productivity. Furthermore, as has already been mentioned, repetition in specific contexts can lead to pragmatic strengthening of (semi-)fixed strings. It is in these and other ways that frequency will be used as an explanatory factor.

By invoking Hilpert's (2013a) concept of constructional change, this usage-based, constructionist view of the progressive can fruitfully be applied to its changing use, offering a novel, comprehensive and more adequate interpretation of the observed phenomena.

The book is structured as follows: Chapter 2 introduces the major assumptions of the usage-based approach to language and of (cognitive) Construction Grammar, including its diachronic application. Chapter 3 summarises previous research on the progressive, addressing the construction's origins, its formal and functional as well as its frequency development. The chapter ends with an evaluation of the state of research and the formulation of the major hypotheses that guide the present study. Chapter 4 introduces the different corpora that are used for the present study and explains the process of data retrieval. It also addresses the different statistical techniques that are employed in chapter 5 and 6. Chapter 5 is concerned with the progressive's frequency development in different genres between the 1950s and today. It focuses on the construction's morphosyntactic and verbal paradigm, highlighting the most important developments behind the progressive's frequency change. Chapter 6 studies (semi-)fixed patterns of use, analysing their development, pragmatic meaning and degree of routinisation. The concluding discussion is provided in chapter 7.

2. Theoretical framework: Usage-based Construction Grammar

This chapter presents the theoretical underpinnings of the present study. First, it is argued that language is best described as a dynamic, usage-based network of choices. Second, Construction Grammar (CxG) is introduced as a theory that regards speakers' linguistic knowledge as a structured network of form-meaning pairings. As will be shown, the usage-based and the CxG approach are highly compatible approaches to the study of language and depend on each other in many respects. It will become clear that frequency effects have systematically been studied from a usage-based perspective, providing essential information for the constructionist approach.

2.1. Usage-based linguistics (UBL)

To begin with, we have to provide a general definition of what is actually meant by the term *Usage-based Linguistics* (UBL). While a focus on language use might ultimately be as old as the study of language itself, a fundamentally usage-based approach to the theory of language, as it is proposed here, is much younger. Ronald Langacker is generally regarded as one of the founding fathers of such an approach. In his *Foundations of Cognitive Grammar*, he describes the usage-based approach to language as follows:

> Substantial importance is given to the actual use of the linguistic system and a speaker's knowledge of this use; the grammar is held responsible for a speaker's knowledge of the full range of linguistic conventions, regardless of whether these conventions can be subsumed under more general statements. [It is a] nonreductive approach to linguistic structure that employs fully articulated schematic networks and emphasizes the importance of low-level schemas. (1987: 494)

A less specific but, arguably, even more pointed definition is given by Kemmer and Barlow: "A usage-based model is one in which the speaker's linguistic system is fundamentally grounded in 'usage events': instances of a speaker's producing and understanding language" (2000: viii). Similar characterisations can be found in Bybee (2010, chapter 1) and Croft and Cruse (2004, chapter 11).

Langacker's definition states some of the most basic assumptions underlying the usage-based model. For example, a speaker's grammar is not conceived of as a streamlined, minimal and highly abstract set of rules, but rather as a large network of linguistic conventions ranging from low to high levels of schematicity. Low-level schemas are

stored in speakers' grammars even though more general schemas exist under which these low-level schemas can be subsumed.

The following sections elaborate on these and other central assumptions of the usage-based model and try to provide the reader with a more or less coherent – but, admittedly, superficial – account of it.

2.1.1. Central concepts and hypotheses

Language is based on domain-general cognitive processes

Arguably *the* most fundamental claim of UBL concerns the cognitive processes underlying language. Unlike the generativist view, which proposes a particular language faculty, the usage-based view sees language as grounded in domain-general cognitive processes, i.e. in processes that also apply to other cognitive domains than language (Bybee 2010: 6ff.; Kemmer and Barlow 2000: xx). In this respect, Langacker states that "I do however subscribe to the general strategy [...] of deriving language structure insofar as possible from [...] general psychological capacities (e.g. perception, memory, categorization), positing inborn language-specific structures only as a last resort" (2000: 2).

Specifically, Langacker (2000: 3ff.) names five psychological processes he regards as essential to language: *entrenchment*, *abstraction*, *categorization*, *composition*, and *association*. Broadly speaking, entrenchment refers to the routinisation and unit formation of linguistic items. Schematization, a special case of abstraction, involves our ability to store and access linguistic structures at different levels of "granularity". Categorization is based on our ability to compare different linguistic structures and to detect similarities and discrepancies between them. Composition "involves the integration of two or more component structures to form a composite structure" (2000: 4). Finally, association is used when one kind of experience is able to evoke another. In language, this is inherent in symbolization, the creation of symbolic relationships.

Bybee, another pioneer in the usage-based field, also names a number of domain-general cognitive processes she considers crucial for the emergence of language (2010: 7). These are similar, however not identical, to the ones proposed by Langacker. Namely, *categorization*, *chunking*, *rich memory storage*, *analogy*, and *cross-modal association*. While categorization and (cross-modal) association have already been mentioned, chunking, rich memory storage and analogy are new. Chunking, however, is closely related to Langacker's notion of entrenchment, which, at its extreme end, leads to the formation of holistic units, i.e. chunks. Rich memory storage refers to "the memory storage of the details of experience with language, including phonetic detail for words and phrases, contexts of use, meanings and inferences associated with utterances" (2010: 7). Bybee further explains that

> Memory for linguistic forms is represented in exemplars, which are built up from tokens of language experience that are deemed to be identical. The primary claim of exemplar representation is that each experience with language has an impact on cognitive representations. (2010: 7f.)

Lastly, for novel utterances to be created based on previously experienced utterances, the process of analogy is invoked (2010: 8).

Reviewing Bybee's (2010) work, Diessel (2011: 832) points out a further cognitive process relevant to the usage-based conception of language: joint attention and the related ability to understand other people's intentions and perspectives. Language as a *social* tool crucially relies on this human ability.

To sum up this brief overview, the proposed list of central cognitive processes involved in language consists of rich memory storage, entrenchment (incl. chunking), analogy, abstraction (incl. schematization), categorization, composition, cross-modal association, and joint attention.[1]

The relationship of synchrony and diachrony

A second fundamental claim of UBL is the inseparability of the synchronic and diachronic analysis of language (cf. Kemmer and Barlow 2000; Bybee 2010, chapter 1). Variation is pervasive in language use, and because usage is put centre stage, it also has to be seen as the locus of language change. Thus, language change is linked to synchronic variation and vice versa.

Importantly, the observed variation unfolds in a structured way, both in the individual's mental grammar and in the community grammar. Kemmer and Barlow put it the following way:

> Patterns in usage data are in general patterns of variation along different dimensions of various kinds, from formal to social. In a cognitive usage-based model, variant linguistic forms can be thought of as alternate possibilities licensed by the linguistic network. The selection of a given entrenched variant for activation is governed by a complex set of motivating factors, including system-internal as well as contextual, situational factors. (2000: xviii)

According to Bybee, variation is closely linked to gradience in the linguistic system – the two phenomena are basically two sides of the same coin:

> Gradience refers to the fact that many categories of language or grammar are difficult to distinguish, usually because change occurs over time in a gradual way, moving an element along a continuum from one category to another. [...] Variation refers to the fact that the units and structures of language exhibit variation in synchronic use, usually along the continuous paths of change that create gradience. (2010: 2)

One important focus of UBL is the processes that create grammatical structures in the course of language change. This focus naturally links UBL to research on grammaticalization (and related processes). The framework of grammaticalization (cf. Hopper and Traugott 2003) is compatible with the domain-general cognitive processes shaping

[1] It is important to note that this list is by no means an exhaustive one. However, the processes stated are certainly *among* the most central ones involved in language.

language. It also explains why grammatical patterns exhibit variation and are gradient, and, finally, helps to account for the lexical specificity of grammar (Diessel 2011: 832).

Thus, the insight that linguistic structure can only be understood in the light of its development is not only compatible with the usage-based account, but constitutes one of its basic assumptions (cf. Diessel 2011: 832).

Exemplar representation and the nature of linguistic categories

In trying to explain the emergence of linguistic categories and constructions from concrete tokens, UBL frequently makes reference to the exemplar model of categorization and learning (Diessel 2016: 222ff.; Ellis et al. 2016: 33ff.; Bybee 2013, Bybee 2010: 14ff.). Diessel explains the model as follows:

> In the exemplar model every piece of information, i.e. every token encountered in experience leaves a trace in memory. Over time, tokens with similar or identical features reinforce each other creating clusters of overlapping tokens known as exemplars [...]. The whole token cluster can be interpreted as an emergent category that functions as an "attractor", i.e. a cognitive reference point for the classification of future tokens. (2016: 222)

Categories based on similar tokens overlap, meaning that there are no clear-cut boundaries between them. Thus, the exemplar model can account for the above-made claim that linguistic categories are gradient and constantly updated and reshaped.

Since memory is cheap and computation costly, the exemplar model is cognitively more plausible than the traditional assumption that linguistic categories are highly abstract entities that have to be matched with concrete tokens (Bybee 2010: 14ff.; Diessel 2011: 834). Cognitive effort is reduced due to rich memory storage.

Exemplar categorization applies at different levels of form and meaning: "phones and their combinations are categorized based on existing representations, as are features of context and meaning" (Bybee 2013: 50). Thus, different categorization criteria are invoked for different types of exemplars. However, the grouping is always based on similarity of the relevant criteria (Bybee 2013: 53f.).

An important determinant of exemplar formation is frequency. As Bybee points out:

> exemplars may differ in strength depending upon the number of tokens that comprise them. That is, exemplars built up from a large number of tokens will be represented more strongly that those built up from a smaller number of tokens. The stronger exemplar or set of exemplars often forms the center of a category and other exemplars are more or less similar to the stronger exemplars or set of exemplars. (2013: 53)

The tokens at the centre of an exemplar cluster represent the most prototypical members of the category (Diessel 2011: 835; Ellis et al. 2016: 34). Consequently, exemplar-based categories also exhibit prototype effects (Bybee 2013: 53), which raises the question of the relationship between prototype theory and the exemplar model. Diessel argues that

the two theories should be regarded as complements rather than alternatives since they focus on different aspects of categorization:

> exemplar theory emphasizes the importance of experience and individual tokens for categorization, whereas prototype theory is concerned with abstract summary representations that are derived from concrete tokens but can license categorization processes independently of them. (2011: 836)

This view is also held by Ellis et al., who state that "human categorization [...] shows effects of both abstract categorical representations and of concrete exemplars" (2016: 60).

Entrenchment and the organisation of linguistic knowledge

I would now like to address a concept that is mentioned in virtually every account of the usage-based model: *entrenchment*. Despite its apparent importance, the concept is seldom consistently defined and often used somewhat vaguely. However, several recent publications have addressed entrenchment in great depth and helped to delimit it more sharply (e.g. Blumenthal-Dramé 2012; Schmid 2015; Schmid 2017b).

Entrenchment makes reference to the formation and organisation of linguistic knowledge. As Schmid (2017a: 9) points out, the term was introduced into linguistics by Langacker, who posits a continuous scale of entrenchment in cognitive organisation. In line with exemplar representation,

> [e]very use of a [linguistic] structure has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact. With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; moreover, units are variably entrenched depending on the frequency of their occurrence. (Langacker 1987: 59; quoted in Schmid 2017a: 9)

Langacker's conception of entrenchment focuses on unit formation – i.e. the gradual process at whose end a deeply entrenched linguistic structure is memorised as a holistic unit. Schmid (2017a: 24), however, argues that entrenchment relates to more than just this one domain-general cognitive process. Apart from holistic storage/chunking, it also commonly refers to the strength of representation of a linguistic structure as well as its degree of schematization. All three aspects of entrenchment depend on frequency of use. Chunking and strength of representation mainly on token frequency, schematization on type frequency (cf. Stefanowitsch and Flach 2017; Blumenthal-Dramé 2017).

In making reference to domain-general cognitive processes, to linguistic variation, and to exemplar representation, entrenchment offers a framework for the formation and organisation of linguistic knowledge that is compatible with all the presented assumptions of UBL. As will be seen in chapter 2.2, entrenchment is also central to Construction Grammar and the conception of constructional networks and links.

It has to be noted, however, that its application to corpus-linguistic results is potentially problematic if no additional experimental data is provided. This is due to the fact

that corpus data alone is hardly sufficient to license definitive statements about certain aspects of entrenchment such as unit formation (cf. Schmitt et al. 2004; Schmid 2010; Dąbrowska 2016: 486ff.). While corpus results can certainly be used to provide *clues* as to whether a specific pattern of use is stored and accessed holistically (e.g. frequency, degree of compositionality, occurrence of hesitations), more definitive insights can only come from psycho- and neurolinguistic experiments that "tap" into the individual's mind much more directly (cf. Blumenthal-Dramé 2012 & 2017 for a psycho- and neurolinguistic perspective on entrenchment). It is exactly because of these possible pitfalls that the present work will use the term entrenchment very cautiously and only to make reference to the strength of representation of a linguistic unit but nothing else (cf. chapter 4.4.1 for discussion).

Priming, alignment and routinisation

The final part of this section is concerned with priming and its consequences for language comprehension, production and change. Priming refers to the "largely non-conscious or automatic tendency to repeat what one has comprehended or produced" (Pickering and Garrod 2017: 173). It works on different levels of language use, for example for words, but also for syntactic structures (Bock 1986 & 2007). Priming effects cannot only be demonstrated in psycholinguistic experiments but also in corpora. For example, Szmrecsanyi (2006) used spoken corpus data to show that speakers re-use recently produced or heard linguistic structures whenever possible – a phenomenon that ultimately goes back to priming effects.

Based on priming, Pickering and Garrod (2004) have come up with a model of alignment in dialogue, which argues that language processing in spoken discourse is very much facilitated through a high proportion of frequently occurring discoursive routines (Pickering and Garrod 2004: 180ff.). Such routines can be stable – i.e. well-established in the language system – or transient – i.e. emerging on the fly in a particular communicative situation (ibid.). Pickering and Garrod (2004) and Garrod and Pickering (2004) link this observation to a process they call *interactive alignment*. Conceiving of conversation as a joint activity (2004: 8), they argue that "in dialogue, the linguistic representations employed by the interlocutors become aligned at many levels, as a result of a largely automatic process. This process greatly simplifies production and comprehension" (Pickering and Garrod 2004: 169). The central mechanism of the alignment process is priming, leading interlocutors to constantly re-use previously encountered linguistic material, resulting in the development of routines during a particular interaction (Pickering and Garrod 2004: 176ff.). These routines make normal dialogue highly repetitive and are understood and produced faster and with less cognitive effort, enabling fluent conversation (Pickering and Garrod 2004: 180ff.).

As Pickering and Garrod point out, "[m]ost discussion of routines focuses on phrases whose status as a routine is pretty stable. Although long-term routines are important, we also claim that routines are set up 'on the fly' during dialogue". Importantly, though, such transient routines can develop into stable, conventionalised routines (Pickering and Garrod 2004: 183), meaning that conversational alignment is a possible starting point

for language change. In this context, interactive alignment constitutes the link between the cognitive mechanism of priming and the socio-cognitive process of *routinisation*, at whose end stand expressions that are fixed to a relatively large degree (Pickering and Garrod 2017: 138). If one acknowledges this connection, one further step – the one to language change – does indeed seem conceivable. Pickering and Garrod (2017) and Mair (2017) discuss this supposed connection from a psycho- and corpus-linguistic perspective.[2]

From a usage-based point of view, Pickering and Garrod's theoretical framework is highly interesting: While priming has an impact on usage, usage is cognitively reflected in priming effects. Through the process of alignment, these can become less transient (short-term routines) and potentially even long lasting (stable routines), resulting in changes in the mental as well as the community grammar. This way, language use shapes the language system.

The next section focuses on an aspect of UBL that is closely connected to the notions of entrenchment and routinisation but that deserves special attention: frequency of use of linguistic structures.

2.1.2. Frequency effects

Usage-based models of language naturally regard frequency of occurrence as an important determinant of linguistic structure and language use (cf. Diessel 2007: 109; Ellis 2012: 7ff.). This is reflected in a number of major publications (for example, Bybee and Hopper 2001; Bybee 2007; Divjak and Gries 2012; Gries and Divjak 2012; Behrens and Pfänder 2016).

Frequency effects are of different kinds and operate on different levels. The most basic and important distinction is that between type and token frequency: token frequency captures how often a certain form (or lemma) occurs (for example in a corpus or in the input of a child); type frequency refers to the number of distinct realisations that occur in an open position of a specific construction (cf. Pfänder and Behrens 2016: 8f.). In the course of this chapter, for example, it has already been mentioned that the strength of exemplar representations is influenced by token frequency. While token frequency also impacts on chunking, it is type frequency that determines the degree of schematization of a linguistic structure.

The following sections present examples of how type and token frequency influence language acquisition, language representation and processing, and language change.

Language acquisition

Usage-based approaches are well-established in the study of child language acquisition (LA) (Ellis 2016: 239) and clearly show that language is not innate but learnt from input (Pfänder et al. 2013: 16). Overviews of frequency effects in LA can be found in

[2] See also Jäger and Rosenbach (2008) and Traugott (2008) for the possible role of priming in grammaticalization.

Ambridge et al. (2015), Diessel (2007), Pfänder et al. (2013), and also Ellis et al. (2016: 45ff.).

Ambridge et al. (2015) argue that frequency effects in LA not only affect single words or fixed strings – where they are least controversial – but also occur on more abstract levels such as inflectional morphology and syntactic constructions. They present evidence for five interrelated theses on frequency effects in LA: First, "Frequency effects exist at all levels and are of many different kinds" (2015: 241). Second, "All other things being equal, frequent forms will be acquired before less-frequent forms" (2015: 242). Third, "High-frequency forms prevent (or at least reduce) errors in contexts in which they are the target" (ibid.). Fourth, "high-frequency forms also cause error in contexts in which a competing, related lower-frequency form is the target" (ibid.). Fifth, "frequency effects will interact with other effects [...] [such as] utterance position" (ibid.).

One particular structure they address in detail is the passive construction. For example, they argue that the learnability problem of the passive in German or English is linked to its low frequency of occurrence. In languages in which it occurs more frequently (for example Sesotho), the passive construction is acquired earlier (2015: 259).

Diessel (2007: 109f.) makes reference to Redington et al. (1998), who studied how children use distributional information to learn the basic English word classes. To do so, they analysed the distributional properties of the 1000 most frequent words in the English CHILDES database. Collecting bigram statistics for each target word, the authors determined how frequent a particular target word was after the 150 most frequent context words. Based on this data, they calculated context vectors that served as numerical representations of the distributional properties of the target words. These context vectors then formed the input of a cluster analysis, which grouped them into lexical classes based on their numerical similarities. The words that occur together in specific clusters closely correspond to the word classes of English. This suggests that "distributional frequencies may play a crucial role in the acquisition of grammatical categories" (Diessel 2007: 110).

Frequency effects are not only attested for L1 but also for L2 acquisition. As Ellis (2012: 10) points out, "there are many commonalities between first and second language acquisition". However, the two processes are different in that

> L2 learners' computations and inductions are often affected by transfer, with L1-tuned expectations and selective attention [...] blinding the acquisition system to aspects of the L2 sample, thus biasing their estimation from naturalistic usage and producing the limited attainment that is typical of adult L2A. Thus L2A is different from L1A in that it involves processes of construction and reconstruction (Ellis 2012: 10f.).

Madlener (2016) shows how type and token frequency can effect L2 learning. Analysing whether learning of the German *sein* + present participle construction is affected by type and token frequency in the input, Madlener conducted a two-week learning experiment with young adult L2 learners of German. The participants were presented with structured input floods in meaning-focused listening comprehension exercises. The results clearly show that

> input flooding in the mere sense of massively increased target token frequencies in the input is clearly not the whole story. As predicted, consistent effects of more fine-grained input features are found, namely in the form of an interaction between overall type frequency and skewed type-token ratios in *structured* input floods. In short, type and token frequency distributions do matter. (Madlener 2016: 166)

While beginning learners profit from moderate type variation with statistical skewing, helping them to identify the construction's function, advanced learners profit from increased type frequency, allowing them to extend the schematic constructional category (Madlener 2016: 152ff.).

Language representation/storage

The role of frequency in language storage/representation has already been mentioned in connection with the exemplar model and entrenchment (chapter 2.1.1). In line with the above-made claims are Croft and Cruse's (2004: 291ff.) assumptions that the independent storage of a word form – regular or irregular – is a function of its token frequency (2004: 293) and that the productivity of a schema is a function of the type frequency of the instances of the schema (2004: 296) (see also Pfänder and Behrens 2016: 9).

In discussing the effects of token and type frequency on exemplar representation, Bybee explains that "exemplars, by their very nature, provide a record of the frequency of occurrence of tokens in linguistic experience" (2013: 59). Like Croft and Cruse, Bybee proposes a direct link between token frequency and the strength of representation. Furthermore, token frequency affects the organisation of categories: "items with higher token frequency within the construction serve as the central members of the categories that form for schematic slots within the construction" (2013: 61).

Finally, as already explained, high token frequency is assumed to facilitate chunking (ibid.). This effect, however, is questioned by Divjak and Caldwell-Harris (2015) and Blumenthal-Dramé (2017), who take a more critical stance than Bybee. As Divjak and Caldwell-Harris explain: "It continues to be debated whether frequency effects are observed because a frequent multimorphemic word or multiword expression is stored as a unit or whether its pieces are more rapidly assembled" (2015: 66f.). Thus, a positive correlation between processing ease and token frequency is not enough to "reject the null hypothesis that higher frequency strings are simply assembled with greater ease and efficiency than matched lower frequency counterparts" (Blumenthal-Dramé 2017: 133). Experiments explicitly testing the relationship between chunking and token frequency in compositional sequences exist but are rare (see Blumenthal-Dramé 2017: 139ff. for an overview).

Regarding the effects of type frequency, increased productivity has already been mentioned. Furthermore, Bybee proposes an inverse relationship between type frequency and semantic coherence (2013: 62). She also argues that highly productive categories often are – but need not necessarily be – highly schematic (ibid.). Importantly, though, Croft and Cruse point out that certain types of a schema/construction that occur with a high

token frequency "may not contribute as much to the entrenchment of the superordinate category schema as instances that are not very entrenched" (2004: 310).

Despite these frequency effects, it needs to be stressed that language representation is much more than a mere tabulation of frequency patterns (Divjak 2012: 3). Divjak and Caldwell-Harris argue that frequency is not the most important factor for creating entrenched representations:

> Frequency is an important contributor, but the relevance of a stimulus for learners' goals may be more important than frequency per se. Entrenchment can occur without repetition frequency, since robust memories can be formed with single-trial learning. [...] But a strong mental representation will be formed from a single instance only in special cases, such as those associated with intense emotions. (2015: 68f.)

Language processing, comprehension and production

As discussed above, increased token frequency may lead to chunking. The corresponding processing phenomenon of chunking is automatization. Diessel explains: "In the psychological literature, the term automatization is primarily used to characterize the way in which controlled processes are transformed into automatic processes through repetition or practice, and the term chunking is primarily used to characterize the way in which automatized sequences are stored and organized in memory" (2016: 229). Thus, co-occurring elements fuse into one and are *stored and processed* as a whole. Apart from the mere token frequency of co-occurring elements/strings, it is transitional probabilities that influence automatic processing, which can be explained by the fact that syntax is sequential in nature and unfolds in real time (cf. Diessel 2016; Pfänder and Behrens 2016: 11ff.).

The resolution of syntactic ambiguities is an area of language comprehension for which frequency effects have been attested. Diessel (2007: 112ff.), for example, reviews evidence for the attachment of prepositional phrases (PP). A PP following a noun phrase (NP) can either be interpreted as being attached to the preceding NP (i.e. [V [NP [PP]]]) or as being attached to the verb phrase (i.e. [V [NP] [PP]]). As Diessel points out, "several studies have found that the attachment site of an ambiguous PP varies with the occurrence of particular lexical items" (2007: 112). He provides the following examples (ibid.):

(2.1) The woman discussed [the dogs [on the beach]].

(2.2) The woman kept [the dogs] [on the beach].

With regard to their surface structure these sentences are identical. However, in an experiment conducted by Ford et al. (1982), 90% of the participants attached the PP to the NP in sentence (2.1), while 95% attached it to the VP in (2.2). Since only the verb varies between sentence (2.1) and (2.2), it must be responsible for the different responses. The difference between *discuss* and *keep* is a semantic one. While the meaning of *discuss*

is the same in both syntactic structures, the one of *keep* is not. Apart from meaning, however, there seems to be another factor at work:

> If we look at the frequency of the two subcategorization frames we find that while *discuss* predominantly occurs with a single NP complement, *keep* tends to occur with two immediate constituents. [...] In other words, *discuss* and *keep* tend to occur in different subcategorization frames, which is eventually motivated by their meanings, but may affect the interpretation of ambiguous PPs as an independent component. Specifically, one might hypothesize that other things being equal people tend to activate the syntactic structure they have encountered most frequently. (Diessel 2007: 112f.)

According to Diessel (2007: 114), the most obvious frequency effect in language production is the phonetic reduction that frequently used expressions tend to undergo (reducing effect) (see also Bybee 2006: 714f.). The effect is due to the fact that high-frequency expressions are highly predictable in spoken discourse. Thus, they are more easily recognised than infrequent structures and, in turn, more likely to be reduced. The phenomenon is particularly evident in linguistic sequences/strings. Diessel explains:

> the contraction of auxiliary verbs (e.g. *I've, he's, we'll*) varies with the string frequency of the subject and the auxiliary. String frequency is defined as the joint frequency of two words, X and Y [...]. [T]he occurrence of contracted auxiliaries varies with the string frequency of a particular subject and auxiliary. Given that pronominal subjects are much more frequent than lexical subjects [...], it does not come as a surprise that auxiliary contraction is largely restricted to pronominal subjects and that the most frequent pronouns are the most common hosts of a clitic. (2007: 115)

Finally, I want to mention recent work by Schneider (2014 & 2016), who shows that chunking of frequently used multiword sequences does not only influence storage and processing but also production in that multiword sequences strongly disfavour interruption by hesitation markers.

Language change

A frequently cited effect of token frequency in language change is one that has just been mentioned in the context of production: phonetic reduction (i.e. the articulatory reduction and coarticulation of frequent expressions) (e.g. Bybee 2010: 37ff. & 2015: 42f.,124f.).

Another effect of token frequency that affects syntagmatic associations is the loss of compositionality (cf. Bybee 2006: 719ff.; De Smet 2017: 88ff.). As syntagmatic sequences undergo chunking, they also often lose their internal constituent structure. Importantly, this seems to be a gradual process rather than an abrupt change (De Smet 2017: 90). A well-known example is the development of *be going to* into a future time expression out of a purpose construction. In the process of grammaticalization, *go* has

lost its original motion sense as the construction underwent reanalysis and acquired a non-compositional meaning that is more than the sum of its parts (Bybee 2006: 721).

Token frequency is also responsible for an effect that runs counter to the above-mentioned reducing effect. It can act as a conservative force for structures that would otherwise be affected be analogical levelling (cf. Bybee 2006: 715,728f.; Bybee 2007: 10f.; Bybee 2010: 66ff.; Pfänder et al. 2013: 20; De Smet 2017: 86f.). Diessel explains:

> frequently used expressions are often resistant to analogical change. For instance, in English there has been continuous pressure to regularize irregular verb forms. Since the time of Old English, nearly 200 verbs have lost the stem vowel alternation and have adopted the regular past tense form [...]. If we look at the verbs that are still irregular in Present Day English, we find that most of them are very frequent. The frequent use has strengthened their representation in memory, which is why they have resisted the pressure from analogical change. (2007: 118)

Semantic bleaching (Bybee 2010: 108), the loss of meaning components and the extension to new contexts of use, is also often considered an effect of token frequency. Repetition is supposed to reduce the effect of the stimulus, resulting in more general usage (Diessel 2007: 117). For example, *a/an* is still singular, but does no longer explicitly specify "one" (Bybee 2010: 108).

De Smet, however, questions this direct link between token frequency and semantic bleaching and regards it as "difficult to find truly convincing examples of bleaching through frequency" (2017: 81). He discusses Bybee's (2003) analysis of the semantic change of English *can* from mental ability via generalized ability to root possibility, in which each step corresponds to the loss of a distinct semantic feature (De Smet 2017: 80). De Smet refers to Coates (1983), who still distinguishes the ability sense (2.3) from the root possibility sense (2.4) for Present-Day English. According to Bybee, though, the changes took place in Early Modern English. In this respect, De Smet explains: "If *can* simply lost semantic specifications, the verb's older more specific sense would have to have been subsumed under the newer generalized sense. The verb would not be felt by speakers to be polysemous" (2017: 81).

(2.3) I *can* walk far mister Brook. I *can* walk all the way to the mine. (Coates 1983: 89, cited in De Smet 2017: 81)

(2.4) We believe that solutions *can* be found which will prove satisfactory. (Coates 1983: 96, cited in De Smet 2017: 81)

(2.5) You *can* start the revels now. (Coates 1983: 88, cited in De Smet 2017: 81)

Furthermore, De Smet argues that the supposed semantic bleaching of *can* is inconsistent with the appearance of the more specific permission sense (2.5) (ibid.). In his opinion, the semantic changes affecting *can* should be seen as a case of pragmatic strengthening (see below) rather than as an example of semantic bleaching (ibid.).

Pragmatic strengthening refers to the process whereby pragmatic inferences become sematicised (De Smet 2017: 79). De Smet explains: "Although there is still discussion about the types of pragmatic inference most typically involved [...], there is no questioning the pervasiveness of the general process in all domains of grammar" (ibid.). Bybee (2007: 17) argues that even though pragmatic strengthening is based on repetition, it does not normally require high frequency. More important is repetition in the appropriate contexts – this alone can cause the formation of a new construction. She provides a well-known example: the *What's X doing Y?* construction (cf. Kay and Fillmore 1999), which expresses incongruity or disapproval as in "What's this fly doing in my soup?". The construction is not particularly frequent; however, its repeated use in the appropriate contexts has resulted in its pragmatic strengthening.

Increasing use of a construction can also lead to innovative uses. Studying the relationship of convention and innovation, De Smet explains:

> There are at least two factors that can shift the odds in favor of the unconventional. First, an unconventional expression will be likely if its deviation from convention is so minimal as to be (almost) undetectable – hence the role of analogy in gradual change. Second, this effect will be the stronger, the more accessible the analogically related conventional expressions are in mental retrieval. This will eventually depend on how well established they are. [...] [T]he more readily retrievable a conventional use of an expression is, the better are its chances of being used also in similar but unconventional ways. (2016: 86)

Thus, increasing token frequency of a construction makes innovative uses more likely. Such innovations can involve new meanings of a word (De Smet analyses the development of the noun *key* into an adjective) but also the use of new verbs in a construction, meaning that token frequency can exert an attraction effect, leading to increased type frequency.[3]

Finally, there is evidence suggesting that token frequency can be part of a positive feedback loop, in which increased frequency leads to even higher frequency. For example, Leech et al. (2009: 269f.) argue that the recent increase of the progressive in 20th-century English might have happened under its own momentum (cf. chapter 3.3.2). Frequent use of the construction makes it mentally more central and accessible, which, in turn, leads to even more frequent use. Similarly, Bybee claims that "items of higher frequency will also be easier to access, which will increase their frequency even more" (2013: 61).

I would like to end this survey of frequency effects in language change by briefly commenting on the role of frequency in grammaticalization. Grammaticalization is one of the most central concepts in the study of language change and is often linked to changing frequency of use. Diessel, for example, states: "Grammaticalization is crucially motivated by semantic (or conceptual) factors. [...] But in addition to the semantic factors, frequency plays an important role in the process of grammaticalization" (2007: 117f.). However, the exact nature of this relationship is often far from clear. In this respect, Mair asks:

[3] Cf. Petré (2016b) for a related approach towards the relationship of conventional and innovative uses.

> Is an increase in discourse frequency a prerequisite for and concomitant of ongoing grammaticalisation [...], or is it a mere epiphenomenon, a post-facto symptom that grammaticalisation has occurred and the newly developed structural option is spreading through genres and styles [...]? (2004: 126)

Mair distinguishes between a dynamic and a static type of grammaticalization. The static type is not linked to changes in frequency and cannot be regarded as a process of language change. The dynamic type refers to changes that are related to frequency shifts. It is this type that is commonly understood by the term grammaticalization. As Mair shows, the dynamic type of grammaticalization "is not accompanied by a simultaneous across-the-board increase in discourse frequency" (2004: 138). Rather, such increases constitute a delayed symptopm of earlier grammaticalization. Importantly, though, it is changes in relative or proportional frequencies that are usually part of the central phase of grammaticalization. The development of *be going to* into a marker of future time serves as an example: "infinitives started crowding out prepositional complements rapidly right during grammaticalisation, and long before the overall frequency of *going to* took off on its dramatic rise" (ibid.).

So far, this chapter has introduced central assumptions of UBL and elaborated on the role of frequency. Putting it in the words of Pfänder and Behrens (2016), it can clearly be stated that "experience counts". However, the effects of frequency are limited, and there are other important factors that interact with them. As Pfänder and Behrens make clear: "The strength and nature of our experience is also influenced by processing factors such as the *context* in which a unit occurs, its perceptual *salience*, and memory related factors such as *recency*" (2016: 6). Ellis stresses the fact that language is not only a cognitive but equally a social phenomenon: "Language usage, social roles, language learning, and conscious experience are all socially situated, negotiated, scaffolded, and guided. They emerge in the dynamic play of social intercourse" (2012: 24). Keeping in mind these considerations is crucial if one wants to avoid conceiving of language as a mere tally of utterances.

2.2. Construction Grammar (CxG): A brief introduction

The second part of this chapter is concerned with the theory of grammatical knowledge pervasive in UBL: Construction Grammar (CxG). First, I will provide a short introduction, offering a definition of CxG and asking in what way UBL and CxG are related or even congruent – addressing, for example, the place of frequency and entrenchment in the CxG framework. Second, I will discuss how CxG influences our understanding of language change.

2.2.1. Introducing CxG

Fundamentals

Construction Grammar has its origins in the outgoing 20th century. Foundational texts include Langacker (1987), Fillmore et al. (1988), and Kay and Fillmore (1999). Goldberg's seminal work on argument structure constructions (1995 & 2006) has exerted a major influence on the more recent CxG discourse.

Construction Grammar assumes that all of a speaker's linguistic knowledge consists of constructions – symbolic units that have both a form and a meaning. Constructions include lexical but also complex phrasal units. They can be very specific (e.g. single words such as *and*) but also very abstract (e.g. the ditransitive construction). Constructions are organised in a highly complex and structured network that is supposed to mirror a speaker's entire knowledge of his or her language: the construct-i-con. In the construct-i-con, constructions at different levels of abstractness are linked vertically by different kinds of inheritance links.[4] Constructions at the same level of abstractness are linked horizontally.[5][6]

Since not only words but also (complex) phrasal and syntactic patterns are seen as constructions, CxG naturally abandons the traditional dichotomy between grammar and lexicon (the so-called *dictionary and grammar model*) and posits a gradual cline instead (cf. Ziem 2014: 17ff.; Hilpert 2014: 3ff.). This comes with the major advantage that idiomatic expressions can naturally be accommodated in the construct-i-con. Idioms are pervasive in language use but have no natural place in the dictionary and grammar model. Hilpert explains:

> many idioms cannot be stored as fixed strings, which makes it necessary to think of idiomatic expressions as schemas with slots that can be filled with certain elements but not others. [...] [T]he patterns that are memorised show characteristics of lexical entries, but also of grammatical rules. (2014: 6)

Abandoning the dictionary and grammar model and positing the existence of a construct-i-con that can handle all sorts of different form-meaning pairings is one of the major characteristics of CxG.

But how exactly are constructions defined? First of all, constructions are primarily *cognitive* entities. This is due to the fact that CxG is designed as a theory of linguistic *knowledge*. Second, constructions are *generalisations* across linguistic patterns encountered in language use (Hilpert 2014: 9). A frequently cited definition is that of Goldberg:

[4] The most basic inheritance link is the *instance link*, which decribes the mapping of a specific construction to a more schematic one. Hilpert (2014: 60) gives the example of the idiom *face the music*. It is a special case of transitive *face*, which is an instance of the transitive construction.

[5] *Subpart links*, for example, link constructions that show partial similarities in form or meaning. However, they do not allow the classification of construction as an instance of the other (Hilpert 2014: 62).

[6] All these aspects are addressed in the following introductions to CxG: Croft and Cruse (2004: 257ff.); Diessel (2015); Goldberg (2003); Goldberg (2013); Hilpert (2014); Hoffmann and Trousdale (2013).

> Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency. (2006: 5)

The definition's last sentence already makes clear that Goldberg's conception of CxG is a usage-based one. Not only non-compositional form-meaning pairings are licensed as constructions, but also fully compositional ones if they are frequent enough. Thus, the form *cars*, while being a fully compositional combination of *car* plus the regular plural morpheme *-s*, is nonetheless likely to have a mental representation of its own due to its high discourse frequency.

Apart from the assumption that a) all linguistic knowledge consists of constructions that are b) abstractions based on language use and c) organised in a complex mental network (the construct-i-con), two further assumptions of CxG have to be stated: First, surface form and semantics are directly linked. Language has no transformational or derivational component (Goldberg 2003: 219; Goldberg 2013: 15). Second, cross-linguistic generalisations go back to general cognitive constraints and to the functions of the respective constructions (ibid.). It is not least these assumptions that set CxG sharply apart from mainstream generativist theories.

Importantly, though, CxG is by no means a completely unified theory of language. While Cognitive CxG (Lakoff 1987; Goldberg 1995 & 2006), Cognitive Grammar (Langacker 1987), and Radical Construction Grammar (Croft 2001) all adopt a usage-based perspective and eschew (complex) formalisation, other CxG frameworks such as Berkeley CxG (e.g. Fillmore 1988; Fillmore et al. 1988) are highly formal and do place less emphasis on language use.[7]

If not explicitly stated otherwise, the present study adopts the framework proposed by Cognitive CxG. An important feature of Cognitive CxG is the assumption that constructions can be of any size, from complex patterns down to inflectional morphemes (Traugott and Trousdale 2013: 5). Perhaps best known is Goldberg's research on argument structure constructions (1995 & 2006). This is outlined by Boas:

> constructions such as the Ditransitive, Caused Motion, or the *Way* construction are capable of supplying a verb's semantics with additional arguments. One of the central arguments for positing meaningful constructions that exist independently of the words that instantiate them stems from the wish to avoid the claim that the syntax and semantics of the clause is projected exclusively from the specifications of the main verb. (Boas 2013: 235f.)

Irrespective of the analysed constructions, though, Cognitive CxG aims at a psychologically plausible model of constructional knowledge by investigating language use and the cognitive principles that structure the network of language-specific constructions (Hoffmann and Trousdale 2013: 8). Thus, Cognitive CxG is not only compatible with the

[7] For an overview of constructional approaches see Croft and Cruse (2004: 265ff.); Traugott and Trousdale (2013: 2ff.); Hoffmann and Trousdale (2013).

assumptions of UBL presented in chapter 2.1 but ultimately based on them. Likewise, the usage-based approach has been informed by constructionist theories. As Diessel explains, "Construction grammar has played an important role in the development of the usage-based approach. In fact, in the literature construction grammar is often described as an integral part of the usage-based approach to the study of grammar" (2015: 298).

Organising the network

One important organising principle of the construct-i-con is the concept of *Motivation* (Boas 2013: 242ff.; Ziem 2014: 21ff.), which assumes constructions to be motivated by extra-linguistic (e.g. bodily) experiences. Relations between constructions are not completely arbitrary but motivated by cognitive mechanisms such as metaphor. Furthermore, the "Scene Encoding Hypothesis" states that prototypical meanings of argument structure constructions are based on prototypical event types, anchored in human cognition (Ziem 2014: 23).

As already stated above, Cognitive CxG explicitly includes frequency into the definition of constructions and their organisation. While it is not entirely clear when exactly a pattern is frequent enough to be considered a construction (Traugott and Trousdale 2013: 5), the "view that regular, but sufficiently frequent expressions are stored in the construct-i-con is not only theoretically viable, but also receives empirical support from psycholinguistic studies" (Hilpert 2014: 14).

Consequently, with regard to how information is stored in the construct-i-con, Cognitive CxG does not adopt the *complete inheritance* model but posits redundant representations instead (cf. Barðdal and Gildea 2015: 23; Boas 2013: 244ff.; Croft and Cruse 2004: 275ff.; Hilpert 2014: 65ff.; Ziem 2014: 26ff.). The complete inheritance model is a parsimonious model, leaving a maximum amount of information to be computed, rather than stored (Hilpert 2014: 65). Information inherited from more abstract constructions is stored only once, and only constructional schemas are represented but not their specific instantiations (ibid.). In comparison, models of constructional knowledge that posit redundant representation[8] assume "multiple memorisations of the same pieces of information across different levels of abstraction. [...] besides general schemas, speakers memorise a great many concrete instantiations of those schemas" (Hilpert 2014: 66). Such models naturally accommodate frequency effects and are also compatible with the concepts of entrenchment and exemplar representation as presented in chapter 2.1.

Hilpert and Diessel (2017) explicitly analyse the role of entrenchment in CxG, asking how it shapes the construct-i-con. They adopt Schmid's (2017a: 24f.) view that it makes sense to regard entrenchment as operating over links between the constructions

[8] One such model is the *full-entry model*, which allows for representation of information at every level of the taxonomic hierarchy of constructions (Croft and Cruse 2004: 276). Goldberg (1995: 97f.) argues that full-entry representation is needed in cases of multiple inheritance, where multiple parent nodes contain contrary specifications of certain constructional properties that cannot be inherited without conflict. Another inheritance model allowing for redundant representation – however not positing that it is preferred in all situations – is the *usage-based model*, "in which patterns of language use are taken as evidence for the independent representation of grammatical information" (Croft and Cruse 2004: 278) (see also Hilpert 2014: 66f).

as opposed to the constructions (i.e. the nodes in the network) themselves (2017: 70f.). They state:

> There is general consensus among usage-based linguists that the co-occurrence of linguistic elements in language use leads to the entrenchment of the associative connections between them in memory, which in turn affects the processing and development of constructions in language acquisition and change. (2017: 60f.)

For example, regarding instance links (i.e. links between abstract schemas and their concrete realisations), "entrenchment facilitates the categorization of concrete utterance tokens, also called constructs [...], into constructions, that is, more general schemas" (Hilpert and Diessel 2017: 61). Less prototypical, i.e. less frequent, instances of a construction are assumed to have weaker instance links to the higher-level construction than prototypical ones, which results in higher processing cost for the listener (ibid.). Other kinds of links, e.g. subpart links, which link constructions that show partial similarities, are also subject to entrenchment: "a construction with strongly entrenched subpart links will be highly transparent, whereas a construction with only weakly entrenched subpart links will appear opaque, so that speakers are more likely to process it holistically" (ibid.). The authors conclude that even though there is not yet a complete theory as to how constructions are connected in speakers' minds, there is certainly evidence that the links of the network are subject to entrenchment (2017: 62).

Since the links are susceptible to change, linguistic knowledge is conceived of in inherently dynamic terms. This view is in line with the assumptions of UBL, which sees language in constant change through language use (cf. Hilpert and Diessel 2017: 70f.). Thus, the construct-i-con is by no means a static but a dynamic, constantly changing network of linguistic knowledge. Consequently, language change has a natural place in CxG. This is exactly what the next section is concerned with: a constructionist approach to language change.

2.2.2. Diachronic CxG

If one took the use of the term *construction* as an indication of CxG's pervasiveness in the field of historical morphosyntax, one would have to conclude that CxG has long been applied to the study of morphosyntactic change, as for example in grammaticalization research. This, however, is not exactly the case. Traugott and Trousdale explain:

> The term 'construction' has been widely used during the past two decades in the literature on morphosyntactic change. It is not always clear what this term is meant to refer to. Usually it is not a form-meaning pairing in the constructionalist sense but rather a phrase or constituent, or the syntactic context in which a grammatical item develops. [...] In general, it can be assumed that authors who use the term 'construction' do not have constructionalist accounts in mind unless they explicitly align themselves with construction grammar. (2013: 31)

This observation also applies to most research on the progressive, which almost never uses the term construction in its strict CxG sense. Nonetheless, much work – especially in the field of grammaticalization – has been concerned with phenomena that do qualify as constructions in the strict sense. One very well-known example is the *be going to* future construction.[9]

Over the past decade, though, several linguists have applied CxG diachronically, which is reflected in a number of substantial publications (for example Bergs and Diewald 2008; Hoffmann and Trousdale 2011; Hilpert 2013a; Traugott and Trousdale 2013; Barðdal et al. 2015). From the perspective of CxG, the most central questions in this context are the following: How do constructions emerge and how do they develop over time? From the perspective of historical linguistics, central questions concern the concept of grammaticalization and whether/how it has to be recast if used in a CxG perspective. This also includes the question how well-established concepts such as reanalysis and extension should be applied in a diachronic CxG framework.[10]

Two terms are frequently used in diachronic CxG: *constructionalization* and *constructional change*. However, depending on the author, they refer to different things. Traugott and Trousdale (2013: 20ff.) and Traugott (2015: 54f.) regard constructionalization as the gradual process whereby new form-meaning pairings are created and added to the linguistic system. The process proceeds through repeated reanalysis and results in changed degrees of schematicity, productivity, and compositionality. Constructional changes are merely viewed as accompanying changes:

> Gradual constructionalization is preceded and followed by a succession of conventionalized incremental steps, which we call constructional changes: A constructional change is a change affecting one internal dimension of a construction. It does not involve the creation of a new node. (Traugott and Trousdale 2013: 26)

Constructional changes are changes in either meaning or form alone (Traugott and Trousdale 2013: 44).

Fried (2013) uses constructional change as a general term and conceives of all construction-related changes as constructionalization, which she defines in close connection to grammaticalization:

> grammaticalization processes are most accurately conceptualized as instances of 'constructionalization': a process that leads to (1) the emergence of a new grammatical pattern (construction) out of previously independent material or (2) a reorganization of an existing construction, leading to an increasingly more opaque meaning of the pattern. (2013: 424)

While the first outcome, the emergence of a new construction, corresponds to Traugott

[9] An early publication explicitly applying the principles of CxG to language change is Israel's (1996) study of the English *way* construction (see Barðdal and Gildea 2015: 8f. for a discussion).

[10] These issues are thoroughly discussed in Barðdal and Gildea (2015), which is an excellent introduction to diachronic CxG.

and Trousdale's (2013) and Traugott's (2015) definition of constructionalization, the second outcome is similar to their notion of constructional change.

Still another approach is adopted by Hilpert (2013a), who uses only the term constructional change, without explicitly distinguishing it from constructionalization. He offers the following definition: "Constructional change selectively seizes a conventionalized form-meaning pair of a language, altering it in terms of its form, its function, any aspect of its frequency, its distribution in the linguistic community, or any combination of these" (2013a: 16).

It is this definition that will be used in the present study to interpret the progressive's development in contemporary spoken English. While Fried's (2013), Traugott's (2015) and Traugott and Trousdale's (2013) definitions are closely linked to grammaticalization processes, Hilpert's definition of constructional change does not only make reference to form and meaning but also to frequency and distribution in the speech community. Thus, it fits most naturally in the usage-based framework. Furthermore, Hilpert's concept of constructional change has been developed and applied in a quantitative, corpus-based paradigm, which makes it particularly suitable for the present study. The remainder of this subchapter will introduce the concept, its delimitation, and its application in more detail.

Hilpert's concept of constructional change

According to Hilpert, constructional change operates on a single symbolic unit and may or may not have an effect on the larger language system (2013a: 16). Changes can occur at the formal or at the functional pole of a construction. Importantly, though, as one pole changes, the other is likely to change as well (2013a: 17). Furthermore, frequency changes also qualify as constructional changes. These can be changes in text (i.e. token) frequency, relative frequency (i.e. affecting the frequencies of constructional variants compared against each other), and also type frequency. As Hilpert points out, changes in relative frequency are particularly characteristic of constructional change (2013a: 12):

> one might think of a polysemous form such as the preposition *over*, which has undergone changes in the relative frequencies of its many senses. Changes in these frequencies will alter the cloud of exemplars that represent the construction in speaker's minds. Even if a change does not create new functions or new structures, a rearrangement of relative frequencies still brings about a constructional change. (2013a: 17)

Constructional change explicitly takes into account the social context by addressing a construction's distribution in the speech community. In this respect, Hilpert explains:

> Even in the absence of any functional, structural, or frequency change, a construction may alter its distribution across different groups of speakers or across different genres. As an example for change in social distribution, quotative *be like* (*And I was like, wow!*) has recently broadened its distribution in Tyneside English [...]. Again, this would constitute constructional change. (2013a: 17)

As depicted so far, constructional change seems to be a very general concept. This raises the question how it compares to other concepts, most importantly grammaticalization and language change in general.

On the one hand, there are many constructional changes that go beyond, or do not qualify as grammaticalization. Most obviously, these are all the kinds of lexical semantic changes that constitute constructional changes (given the view that lexical items are also regarded as constructions) (2013a: 9). But there are also grammatical changes that fulfil the criteria of constructional change but not of grammaticalization. One such case would be the loss of the verb-second (V2) constraint in the history of English. While this change is not normally regarded as grammaticalization, it does, however, qualify as constructional change: "From the perspective of Construction Grammar, the loss of V2 could be seen as the deterioration of a general syntactic construction" (2013a: 10).

Furthermore, many frequency changes, which constitute constructional changes, are not necessarily part of grammaticalization processes (2013a: 12). Mair (2004) has shown that the relation of grammaticalization and frequency changes can be quite complex. Regarding the comparison of constructional change and grammaticalization, the important point is that "frequency changes that have been discussed as concomitants of grammaticalization do not exhaust the spectrum of frequency changes that can be observed in the developments that constructions undergo" (Hilpert 2013a: 12).

On the other hand, there are changes that qualify as grammaticalization but not as constructional change, meaning that the latter does not simply subsume the former concept (ibid.). The formation of grammatical paradigms beyond the level of a single construction is a case in point: Hilpert makes reference to Lehmann (1995) and his concept of paradigmatization, i.e. the tendency of grammaticalizing constructions to form paradigms or to become integrated into existing paradigms (2013a: 11). The emergence of the English modal auxiliaries is an example of such a process. Theoretically, one could posit the existence of a general modal auxiliary construction comprising all existing formal realisations. However, Hilpert doubts "whether such generalizations are still to be regarded as constructions, that is, as signs. [...] [N]ot all linguistic generalizations are, simply by virtue of being generalizations, also constructions" (ibid.).

Finally, Hilpert asks how constructional change is different from the general notion of language change. As an example, he uses the Great Vowel Shift, which affected different words indiscriminately of their grammatical status. Thus, like many other language changes, and like emerging paradigms in grammaticalization, the changes that brought about the Great Vowel shift exceeded the level of individual constructions and, consequently, do not qualify as constructional changes (2013a: 13ff.). This means that "not all change in language is necessarily constructional" (2013a: 16).

Apart from these theoretical considerations, it is worth asking how constructional changes can best be studied. Hilpert adopts a quantitative corpus-based approach, in which questions are formulated in such a way that corpus frequencies can provide answers to a wide range of questions concerning form, meaning, and distribution of a construction. Making reference to exemplar representation, he explains:

> The idea that constructions are mentally represented as clouds of exemplars

> [...] is helpful in this regard. The instances of an exemplar cloud will vary along multiple dimensions. This variation can be assessed on the basis of corpus data, and changes in this variation can be tracked diachronically. One variant of a construction, that is, one subtype of the construction that has a certain configuration of features, may become more frequent over time, as other variants become less frequent. Entirely new variants may develop. One constructional subtype may develop new extensions, other subtypes may converge on a single form or function. (2013a: 6)

Such changes are likely to happen along many dimensions at the same time, which means that the identification of a constructional change can be difficult. It may require fairly sophisticated statistical tools in combination with in-depth qualitative analysis (2013a: 6ff).[11]

Hilpert (2013a) presents different case studies to test and develop his theoretical framework. He researches constructional change in allomorphy, word formation, and syntax. One of his case studies analyses the development of the so called *V-ment* construction – normally referred to as the nominalisation suffix *-ment* (2013a: 115). He defines the construction as follows:

> The *V-ment* construction consists of an abstract slot for a lexical stem and a suffix with the phonemic structure [mənt]. The stem strongly tends to be verbal, but exceptions with nominal or adjectival stems are attested. In its instantiations, the two parts of the construction form a noun that typically conveys the meaning of an action (as in *adjustment*), the results of an action (*assortment*), or the means to accomplish an action (*refreshment*). (2013a: 115)

The construction originated in the Middle English period, as a consequence of borrowing from French (e.g. *judgement* or *payment*). In the course of the fourteenth century, a productive schema evolved. Ever since, however, the construction has continually declined in productivity and is no longer productive in Present-Day English (2013a: 113). Hilpert analyses how this development "unfolded in terms of its quantitative dynamics" and what "qualitative changes [...] the *V-ment* construction underwent during its short but eventful life" (2013a: 114).

After determining developmental stages in the construction's history (by applying a variability-based neighbour clustering algorithm to the values of the construction's (expanding) productivity) (2013a: 133f.), Hilpert identifies five more variables that he considers important with regard to the construction's diachronic development. These are source (borrowed vs. derived), stem type (verb, adjective, noun), branching (binary, left, right), transitivity (transitive, intransitive), and semantic type (activity, result, means, place) (2013a: 134ff.). Using Hierarchical Configural Frequency Analysis (HCFA)[12], Hilpert is able to show how complex, multivariate interactions between the different

[11] For an introduction to some of these statistical procedures see Hilpert (2013a, chapter 2) and Hilpert (2013b).

[12] See also chapter 4.4.4.

variables play out at different diachronic stages, revealing which constructional usage patterns were typical/atypical (i.e. more or less frequent than statistically expected) in different periods. Adopting such a multivariate approach, one can find overarching trends, while also being able to detect lower-level trends that would otherwise go unnoticed (2013a: 148).

The results clearly show that "the development of the English *V-ment* construction is more than merely a story of rise and fall in the domain of morphological productivity" (2013a: 152). Hilpert finds a very complex set of different developments that can only be accommodated under a constructional perspective (ibid.). The construction first emerged as a pattern borrowed from French. Host elements were transitive verbs. However, between 1250 and 1400 the use of the construction was very diverse, open to different host classes and semantic interpretations. An adjectivally based type emerged for a short period, remained negligible, though. Now, a crucial developmental stage was reached:

> From the fifteenth century onward, a type of the *V-ment* construction establishes itself that subsequently comes to dominate the development of the construction. Forms with a native, transitive, internally complex verbal stem and an action interpretation represent a pattern that continues to be highly successful throughout the nineteenth century, and that only fades as the *-ment* suffix as such ceases to be productive (2013a: 153).

Thus, the construction's development is a diverse one, defying one unidirectional interpretation. Different patterns of use, or constructional subschemas, are productive at different developmental periods. This leads Hilpert to conclude that "it is descriptively inadequate to think of the *V-ment* construction as a single word formation process for which a common measure of productivity could be computed" (2013a: 154). The major strength of a multivariate, quantitative approach lies in the fact that it can "empirically determine the structural and semantic characteristics of these subschemas, their respective level of abstraction, as well as the time window during which they were most productive" (ibid.).

2.3. Summary and relevance for the present study

This chapter has introduced a usage-based conception of language and the concept of Construction Grammar. Major assumptions of both frameworks have been explained, and it has been argued that UBL and CxG are highly compatible. Finally, the diachronic application of usage-based CxG has been addressed – especially in the context of Hilpert's concept of constructional change.

With regard to the usage-based approach, a major topic of this chapter were frequency effects. These have been reported for language acquisition, representation, processing, and change. It has also become clear that frequency is a crucial factor in CxG. While not all of the studies reported in chapter 2.1.2 have a clear-cut constructionist focus, their results are nonetheless applicable to the CxG framework. Usage-based approaches

to CxG naturally allow for frequency effects of different kinds at different levels of the construct-i-con. As already pointed out, frequency is even part of Goldberg's well-known definition of a construction.

While this insight is generally acknowledged in usage-based CxG approaches, it has – naturally – received different levels of attention from different authors. Traugott and Trousdale's (2013) work on diachronic CxG, for example, represents a rather classic philological approach, in which qualitative analysis of corpus data figures prominently. On the other hand, work by Bybee (e.g. 2013), Hilpert (e.g. 2013a & 2014), and Ellis et al. (2016) puts a strong focus on concepts such as exemplar representation, entrenchment, and frequency effects. The following statement by Bybee serves as an example:

> both type and token frequency are important to our understanding of constructions as they affect category formation for slots in constructions, productivity of these slots, as well as the degrees of analyzability of constructions and particular exemplars of constructions. As frequency representations are an integral part of exemplar representation, these facts support exemplar models as a good choice for Construction Grammars (2013: 63).

Recent research has made use of a range of frequency-based measures that offer very fine-grained insights into the use and potentially also the cognitive representation of constructions (for overviews see Gries 2013a; Gries and Ellis 2015; Stefanowitsch and Flach 2017). Perhaps the best-known example is collostructional analysis, which comprises three different methods that compute a bidirectional association measure between constructions and words (Gries 2013a; Stefanowitsch 2013).[13]

Hilpert's concept of constructional change regards frequency as a key factor and is compatible with the use of statistical measures to gauge the exact impact of different kinds of frequency changes.

All this shows that the study of frequencies and frequency effects, as well as the application of frequency-based measures have a natural place in usage-based approaches to CxG.

With regard to the recent use and development of the progressive, such a usage-based CxG account that puts a strong focus on frequencies and their statistical analysis promises to be highly productive. As will become apparent in the following chapter, previous research on the progressive has rarely adopted an explicit CxG perspective. Furthermore, many accounts undeniably have a strong focus on actual usage – by investigating corpus data – but do not make extensive reference to concepts such as entrenchment or routinisation, frequency effects, or exemplar representation. The present study aims to do exactly this, and asks whether the perceived unruliness of the progressive can be better accounted for in a usage-based CxG framework. How can prototypical aspectual uses and rarer, pragmatically motivated idiomatic uses be integrated in one constructional network? Which developments does such a network undergo at different levels of generality? Is the progressive's changing use really just one change, or rather a complex development consisting of several changes at different levels and times (like

[13] See chapter 4.4.3 for a detailed account.

the development of the *V-ment* construction)? And finally: Can frequency effects be observed? Is higher token frequency accompanied by increasing type frequency, or is the frequency change taking place under its own momentum? What is the role of frequency regarding the use and development of lexically specific usage patterns?

It is these and other questions that the present study seeks to answer. Hopefully, it has become clear to the reader that a usage-based CxG approach offers a very promising starting point for this endeavour.

3. The progressive: Survey of previous studies

3.1. Defining the progressive

The English progressive construction is formed by the auxiliary *be* and a following *-ing* participle and is generally considered a marker of *aspect*. Comrie defines aspects as "different ways of viewing the internal temporal constituency of a situation" (1976: 3). Similarly, Brinton refers to aspect as "a matter of the speaker's viewpoint or perspective on a situation" (1988: 3). According to Brinton (ibid.), two rather different phenomena are often subsumed under the label of aspect; namely grammatical and lexical aspect. The former is covered by the above definitions and can be regarded as aspect in the strict sense. The latter, which is also referred to as *Aktionsart*, does not concern the speaker's point of view, "but the inherent nature of the situation portrayed" (ibid.).[1] Brinton explains that:

> Aspect is grammatical because, broadly speaking, it is expressed by verbal inflectional morphology and periphrases, aktionsart by the lexical meaning of verbs and verbal derivational morphology. Aspect is subjective because the speaker chooses a particular viewpoint, whereas aktionsart, since it concerns the given nature of the event and not the perspective of the speaker, is objective. [...] [T]he aspectual interpretation of a sentence depends on an interaction between these two categories.[2] (1988: 3)

The progressive construction encodes grammatical aspect and is typically considered a subtype of the *imperfective*,[3] which treats a situation as being unbounded in time (cf. Comrie 1995: 1244). However, progressive meaning differs from imperfective meaning in that it combines "(non-Habitual) Imperfective aspect with dynamic (as opposed to stative) semantics" (Comrie 1995: 1245).

As Brinton points out, the progressive is both contrastive and obligatory in English (1988: 6). As fully grammaticalized aspect marker (formal category) it is generally required to express progressive aspectuality (semantic category) (see also Huddleston and Pullum 2002: 162f.). However, compared to progressive constructions in other languages, the English progressive has a very wide meaning range that goes beyond the core

[1] See Vendler (1957) for a classic treatment of lexical aspect in English, and Croft (2012:33ff.) for a critical discussion thereof.

[2] See also Kortmann (1991: 13), who states that "Aktionsart [...] has nothing to do with grammar but relates solely to the semantics of verbs and predicates [...]."

[3] See Dahl (1985: 92f.) for a critical discussion of this claim.

aspectual meaning (cf. Comrie 1976: 37ff.; Kranich 2013: 4ff.).[4] Habitual progressives, so-called "subjective" uses and also progressives expressing future time reference do not readily fit the notion of progressiveness. Also the aspectual meaning itself is far from being uncontroversial. For example, some authors assume the progressive's aspectual meaning to be developing from progressive to imperfective (cf. Comrie 1995: 1245), along the lines of a frequently observed diachronic trend (Bybee 2015: 144f.).[5]

The following section will briefly address the progressive's origins and delineate its development leading up to the 20th century.

3.2. Origins and development prior to 1900

3.2.1. Source(s) of the progressive

Even though the topic has received a fair amount of scholarly attention (e.g. Kranich 2010; De Groot 2007; Smith 2007; Scheffer 1975), the progressive's linguistic source is still a matter of controversy. Two Old English (OE) constructions are discussed as possible origins of the modern progressive:

1. A participial construction of the type *beon/wesan* + present participle: e.g. *he was blissiende* (Kranich 2010: 80).

2. A locative prepositional construction of the type *beon/wesan* + preposition + gerund: e.g. *he was an hontyng* (Smith 2007: 206).

Scheffer (1975: 245) proposes that the Modern English (ModE) progressive can be seen as a direct continuation of the OE participial construction.[6] Núñez-Pertejo (2004: 112) calls this view the "theory of continuation" and notes that most recent work on the topic seems to adopt it. So for example Kranich, who describes the early development of the participle construction as follows:

> primary grammaticalization by which the combination reached the status of a periphrastic construction, had already taken place within the OE period. [...] [T]he function of this construction was, however, not yet clearly grammatical. In OE and ME, it is only for the sake of a uniform denomination that we speak of the "progressive", since the construction was in general not yet used to express progressive aspect. (2010: 249)

However, as Kranich (ibid.) explains, one of the functions of this early "progressive" was to emphasize the durative and/or imperfective nature of the situation; a function that would later develop into progressive aspectuality.

[4] See also König (1995b) and Klein (1995) for two concise discussions of aspectual and verbal distinctions in English.

[5] See Kranich (2010: 32ff. & 2013: 22ff.) for a critical discussion of this claim.

[6] The participial ending *-ing* became substituted for *-end(e)* in Early Middle English (De Groot 2007: 176).

Counter to this theory of continuation runs the "prepositional theory" (Núñez Pertejo 2004: 113ff.), which sees the much rarer prepositional construction as the forerunner of the present-day progressive. As Núñez-Pertejo explains, proponents of the prepositional theory

> put special emphasis on the functional dissimilarities that exist between Old English and modern English periphrases to justify that Middle English *be + -ing* could have hardly developed out of Old English *beon/wesan + -ende*. Therefore, the origins of the cluster should be searched elsewhere. According to this theory, prepositional patterns of the type *be on hunting* were the real ancestors of *be + -ing* and not participial phrases of the type *beon/wesan + -ende*, and, therefore, the semantics of the modern progressive would show its locative origin, which Present-day English *be + -ing* has acquired from the prepositional phrase. (2004: 116)

Bybee et al. (1994: 132) stress the point that the majority of progressive constructions in other languages can be traced back to locative sources, which would suggest a locative source for the English progressive as well. Similarly, K. Aaron Smith suggests a "strong universal relationship between locativity and progressivity" (2007: 230).

However, it has to be pointed out that the locative prepositional construction was very infrequent not only in OE but also throughout the Middle English (ME) period. (cf. Kranich 2010: 78; Núñez Pertejo 2004: 116; Scheffer 1975: 244ff.) What is more, it seems to have occurred with a very restricted set of verbs – such as *hunting* and *riding* (De Groot 2007: 186) –, which Kranich regards as an indication of a low degree of grammaticalization "since the expression of activities such as hunting is still easily connected to the original locative meaning of the construction" (2010: 78).

This strong connection with the original locative meaning is interpreted as an *absentive* function by De Groot (2007). He regards the participial construction as the actual forerunner of the progressive and argues that the "construction with the verbal noun *[i.e. the prepositional construction; U.R.]* was not a (kind of) progressive too" (2007: 187). However, he acknowledges the semantic relation between absentive and progressive and assumes that the two constructions might have fused into one in ME, as the absentive was reinterpreted as a progressive (2007: 187f.) Such a coalescence of the two constructions is what Núñez Pertejo (2004: 117f.) refers to as the "amalgamation theory". It is also advocated by Killie (2008: 84f.), who considers the progressive a blend of at least these two sources.[7]

Regardless of the precise origin of the progressive, the construction soon began to establish a firm hold in the English language. In early ME its frequency was still quite low, but it had already doubled by the end of the ME period; heralding what Fischer (1992: 251) calls an "astronomical" frequency increase from the beginning of the ModE period onwards. The following section will address these frequency developments in more detail.

[7] Unlike De Groot, however, Killie interprets the participial construction as an expressive device and the prepositional construction as the one conveying aspectual meaning.

3.2.2. Frequency development

A wealth of publications has addressed the progressive's frequency development. Since it is beyond the scope of the present chapter to do justice to every single one of these, I decided to focus on only those I consider the most instructive ones.

Fischer (1992) gives an overview of the progressive's frequency development in Old and Middle English. She presents figures by Nehls (1974: 139), indicating that the progressive (i.e. the participial construction) reached considerably higher frequency levels in OE than in early ME. The rate recorded for early ME is as low as six tokens per 100,000 words. By the end of the period, though, the frequency had roughly doubled (Fischer 1992: 251). Fischer considers the "enormous expansion of the functional load of the form in *-ing*" (1992: 252) – a result of the confusion of the inflectional ending of the present participle and the derivational morpheme creating verbal nouns – as one explanatory factor for this frequency increase in late ME. The resulting coalescence of the prepositional and the participial construction "sharply increas[ed] the frequency of the progressive form proper" (1992: 253).

Núñez Pertejo (2004: 99) provides a similar account of the progressive's development in ME and states that "at least in early Middle English, the frequency of occurrence of *be + -ing* was lower than the frequency of occurrence of Old English *beon/wesan + -ende*". She refers to Mustanoja (1960: 585f.), who claims that higher Northern usage of the progressive spread over the country in the course of the 14th and 15th centuries, leading to much higher frequencies in all dialectal areas by the end of the ME period.

Núñez Pertejo (2004: 160ff.) also analyses data from the Helsinki Corpus to provide a clearer picture of the progressive's development and use in Early Modern English (EModE). Even though her total number of progressives is quite low (N=178), she shows that frequencies increased steadily and significantly during EModE. For the first period from 1500 to 1570 the so-called M-coefficient (i.e. the frequency per 100,000 words) is 16.82, for 1570 to 1640 it increases to 26.87 and finally reaches a level of 55.54 for the period from 1640 to 1710. Compared to frequencies in present-day English (PDE) these figures are still low, which leads Núñez Pertejo to conclude that "the progressive was not a frequent construction in early Modern English" (2004: 161).

Analysing the progressive's distribution across different genres, Núñez Pertejo finds that "*be + -ing* periphrases are, as a general rule, more frequently attested in informal writing" and occur more frequently "in genres closer to actual speech than in other written genres" (2004: 173f.).

Rissanen also addresses the progressive's frequency development in EModE and argues that "the seventeenth century is the crucial period in the development of the progressive" (2000: 216). He refers to Elsness (1994), who, like Núñez Pertejo, studied the progressive in the Helsinki Corpus and also finds a roughly threefold increase in absolute terms.[8]

Denison reviews the progressive's frequency development in Late Modern English (LModE) and states that "[t]he progressive construction [...] has undergone some of the most striking syntactic changes of the lModE period. By early in the ModE period the *BE + -ing* pattern was already well established, and its overall frequency has

[8]Elsness's absolute numbers (1994: 11) differ only very marginally from Núñez Pertejo's (2004: 160).

increased continuously ever since" (1998: 143). Specifically, he refers to Dennis (1940) and Arnaud (1983). The former "estimates an approximate doubling every century from 1500", while the latter – analysing a private letter corpus – "estimates a threefold increase during the nineteenth century alone" (Denison 1998: 143).

Scheffer (1975: 250ff.) presents a list of literary works from the 16th to the 18th centuries, which is adapted from Mossé (1938), and which shows that the M-coefficients for works of prose have increased substantially over time. For poetry the figures are considerably lower and no obvious trend can be observed. Scheffer also provides a list of progressive frequencies for dramas; they are, however, not normalised and consequently not easily interpreted. However, since the absolute numbers clearly increase diachronically, one can assume that progressive use in dramas became more common between the 16th and 18th centuries. Finally, based on a list of works from the 19th and early 20th centuries[9], Scheffer also infers increasing progressive use for this period.

A frequently cited study is Strang (1982). Like Denison (1998), she draws on figures from Dennis (1940) and concludes that

> the pattern is familiar enough by early Modern English, and goes on getting commoner very fast. The fullest attempt at quantification puts the rate tentatively at a doubling every century since 1500 to the present day, with a flattening of the curve in the eighteenth century made good by a spurt at the beginning of the nineteenth. (1982: 429)

Furthermore, Strang (1982: 430ff.) presents figures from a corpus of novels spanning from the early 18th to the 20th centuries. For 18th-century novels her M-coefficient is 107, increasing to M=243 in the 19th century and finally reaching the surprisingly high number of M=657 for 20th-century novels (1982: 464). Even though it is not entirely clear how she arrives at this number, Strang (1982: 432) suggests a rate of M $\approx$ 265 to M $\approx$ 330 as "representing the maturity of the construction". Thus, frequency-wise the progressive would have reached full maturity in the course of the 19th century.

Hundt (2004a & 2004b) provides numbers from ARCHER[10], spanning a period of about 350 years from 1650 to the 1990s.[11] With one exception (1850-99 till 1900-49) progressive frequencies increase continuously (2004a: 58). Between the periods of 1800-49 and 1850-99 a particularly pronounced frequency increase can be observed (from M=101 to M=196).

Kranich's (2010) dissertation on the progressive in Modern English is without doubt one of the most detailed and comprehensive studies on the construction's historical development. Using the British component of ARCHER-2[12], she covers a time span from

[9] Again, the list is taken from Mossé (1938).

[10] *A Representative Corpus of Historical English Registers.*

[11] The figures presented in Hundt (2004b) are slightly higher than those in Hundt (2004a), even though the same corpus was used. This is due to the fact that in Hundt (2004b) the so-called *going to*-future was not excluded from the counts, whereas in Hundt (2004a) it was. I refer to the figures of Hundt (2004a).

[12] ARCHER-2 is an extended version of the ARCHER corpus, comprising approximately 1.36 million words.

Time span	M-coefficient
1600-1649	32
1650-1699	61
1700-1749	84
1750-1799	103
1800-1849	139
1850-1899	243
1900-1949	350
1950-1999	393

Table 3.1.: Frequency increase of the progressive in ARCHER-2 (Table 3b in Kranich 2010: 95). The M-coefficient indicates the progressive's frequency per 100,000 words.

1600 to 1999 and a variety of different genres, yielding a total of 2,662 progressive tokens (2010: 10f.). Similar to Hundt (2004a), she reports a continuous frequency increase. Her findings are summed up in Table 3.1.

The numbers are similar to those of Hundt (2004a: 69), with exception of the first half of the 20th century. For this period Kranich reports a much higher progressive frequency than Hundt, which can only be due to different sampling procedures and the different versions of ARCHER. Judging from Kranich's numbers, the second half of the 19th (and the first half of the 20th) century was a period of rapid growth.

Regarding progressive frequencies in different genres, Kranich shows that colloquial and/or speech-based genres clearly favour the use of the progressive. Particularly high numbers are reported for Drama and Letters, while lowest numbers are obtained for Science and Medical writing (2010: 101). For the early periods progressive frequencies in Fiction are lower than those in Drama and Letters, but from the second half of the 18th century onwards, numbers pick up and reach a similarly high frequency level. According to Kranich, this may point to the fact that "the progressive was at first more associated with spoken, colloquial use [...] and that it only took ground in a less speech-based genre such as the novel somewhat later" (2010: 102).

For Journals and News, Kranich's numbers never depart far from the average numbers, meaning that these two genres "neither favour nor disfavour the use of the progressive" (2010: 102). With the exception of one subperiod, the same holds true for the speech-based but conservative genre of Religious Sermons.

Finally, I would like to address Smitterberg's study on the progressive in 19th-century English (2005). Using the one million word CONCE[13] corpus, he provides a thorough analysis of many different aspects of progressive use between 1800 and 1900. Like Hundt (2004a) and Kranich (2010) he reports a clear frequency increase. For period one (1800-30) the M-coefficient reaches a level of 172, for period two (1850-70) it is M=263, and

[13] *Corpus of Nineteenth-Century English.*

for period three (1870-1900) it is M=316 (2005: 60).[14] Smitterberg links this rise in frequency to his concept of integration, which comprises developments such as grammaticalization, obligatorification, an expanding formal paradigm, and increasing frequencies (2005: 57f.). Consequently, he takes the observed frequency change as an indication that "the progressive became more fully integrated into English grammar during this period" (2005: 58).

The pace of the overall increase reported in Smitterberg (2005) is relatively even, which is, however, not the case for the development in the individual genres (2005: 61ff.). Smitterberg observes that progressive frequencies are highest in the spoken/speech-based genres Trials and Drama, as well as in the informal Letters genre (2005: 63), which is in line with Kranich's (2010) findings. Lowest frequencies are found for Science and Debates. While the low frequencies for Science are again in line with Kranich's findings, the low numbers in Debates are somehow surprising given the fact that Debates belong to the spoken modality. Smitterberg links it to their formal character and to transcription conventions at the time (2005: 64).

The three genres exhibiting lowest progressive frequencies – Debates, History and Science – are classified by Smitterberg as "expository" genres. To the other four genres – Drama, Fiction, Letters and Trials – he refers as "non-expository" genres. According to him, high progressive frequencies are characteristic of non-expository genres (2005: 65ff.). This observation holds over all three subperiods: "In none of the periods is there any overlap of M-coefficients between these two genre groups" (2005: 67). Importantly, the overall frequency differences between the genres increase rather than decrease in the course of the 19th century (ibid.). For example, a noticeable frequency increase occurred in Science writing, but the increase in Drama was even more pronounced, leading to a larger frequency gap between the two genres at the end of the century. Altogether, this leads Smitterberg to conclude that "the progressive is an oral rather than a literate feature, as the construction is decidedly more common in popular than in specialized genres" (ibid.).

The following sections will address the major functional and formal developments that have occurred together with the construction's frequency increase.

[14] Using a different frequency metric, the so-called S-coefficient, the frequency increase gets less pronounced. As Smitterberg points out: "the S-coefficient increased by 71% between periods 1 and 3, whereas the M-coefficient increased by 81%" (2005: 62). This is due to the fact that unlike the M-coefficient, the S-coefficient is calculated relative to the number of verbs in the respective subcorpus, minus a number of verbal contexts (e.g. imperatives) in which the progressive cannot normally occur. Thus, from a variationist viewpoint, the S-coefficient is the more accurate metric than the M-coefficient (2005: 45ff.). If, for example, a language gets more "verby" over time, i.e. contains more verbs per a certain number of words, this is reflected in the S- but not in the M-coefficient (2005: 62) (see chapter 4.4.1 for a discussion of different frequency metrics).

3.2.3. Functional and formal development

Grammaticalization as aspect marker

The progressive's crucial functional development was its grammaticalization as a marker of aspect. Kranich (2013) refers to this development as the progressive's *secondary grammaticalization*. She explains:

> Secondary grammaticalization is understood as the process by which grammatical forms and constructions become more grammatical, e.g., by acquiring more clearly grammatical meanings. Primary grammaticalization, by contrast, is understood as the process by which lexical elements first become grammatical [...]. With regard to the progressive, primary grammaticalization had already taken place by OE times, when the progressive must have emerged as a construction [...], while secondary grammaticalization can be said to have occurred mainly in the course of the ModE period, during which the function of the progressive construction became clearly aspectual. (2013: 22, footnote 14)

As the root of the progressive's current aspectual meaning Kranich identifies an *imperfective-durative* use as in the following example:

(3.1) *Helas, oh, how heavy a yoke he laid on all who on his time* ***were living*** *on earth.* (Boethius, 39/31-33, Kranich 2013: 22)

However, the aspectual meaning was by no means the predominant one in OE and ME, where progressive use was frequently motivated by a speaker's subjective evaluation of a situation as remarkable, dramatic, or worthy of a vivid description (Kranich 2010: 88). Kranich refers to these uses as subjective uses and explains that only as late as in the EModE period did the aspectual uses become more frequent than the subjective ones (Kranich 2013: 23). In OE and ME, however, "the progressive is never obligatory and has not yet acquired a clear grammatical function" (Kranich 2010: 89). Fischer (1992: 254) points out that the progressive has to be seen as a predominantly stylistic device before it became fully grammaticalized as a marker of aspect.

Killie (2008) provides a detailed analysis of the progressive's semantics in the OE, ME and EModE period. She shows that OE and ME progressive use is characterised by very high relative frequencies of subjective progressives. So-called narrative progressives (emphatic meaning, used to mark peaks in a narrative) account for 21% of all progressives in OE, but go out of use in ME and EModE. So-called stative progressives (providing emphasis, referring to facts, relations, feelings, etc.) are common in OE (22% of all tokens) and even more so in ME (56%/42% of all tokens, depending on the corpus). In EModE their numbers decline (19% of all tokens). Thus, in OE, 43% of all progressives analysed by Killie express a subjective, non-aspectual meaning, while in EModE it is only 21% (2008: 80ff., Table 2&3). This decline in subjective usage is linked to an increase in aspectual usage: so-called focalised uses (presenting an event as dynamic/ongoing at a point in time) clearly increase over time,[15] which "may be seen as a result of the increased

[15] 28% of all tokens in OE, 19%/46% (depending on the choice of corpus) in ME, 61% in EModE.

grammaticalization of the English progressive as an aspectual, focalizing device" (Killie 2008: 79).

Petré (2016a: 43) observes stable progressive frequencies in ME, but shows that focalised contexts of use did nonetheless increase their share at the expense of stative uses. He argues that "towards the end of the Middle English period, ongoingness became semanticized as a component of the lexical meaning of BE V*ing*." (2016a: 33). This means that critical changes in the progressive's development towards an aspect marker were already on their way before 1500.[16]

According to Strang, however, "the rules for the use of the construction were established in the seventeenth century" (1982: 429). She classifies the time before 1600 as a time of unsystematic use and the time after 1700 as one of systematic/grammatically required use (ibid.).

Kranich shows that in the second half of the 17th century 80% of all progressives in ARCHER-2 express an aspectual meaning. It was then and in the early 18th century that the break-through of the aspectual use of the construction occurred and that it became grammaticalized as a marker of progressive aspect (Kranich 2010: 240f.; Kranich 2013: 23).

Smitterberg attests that around 1800 the progressive's functionality was already very similar to the one of today and that "over the 19th century, the paradigm was to become identical to that of Present-Day English, and the progressive was to become near-obligatory to express ongoing action and frame-time in non-stative situations" (2005: 54).

It is obvious that the construction's secondary grammaticalization as aspect marker must have contributed substantially to the frequency increase between 1700 and 1900, because the use of the progressive was more frequently required grammatically. No less important, however, is the development of the construction's formal properties, which is addressed in the following section.

The development of the progressive's formal paradigm

The progressive perfect and pluperfect were first recorded in the second half of the fourteenth century but remained sporadic till the ModE period (Fischer 1992: 255f.). In EModE present and past progressives are the most common forms, but many other forms are used already. These are perfect and pluperfect progressives, progressives with modal auxiliaries and non-finite constructions such as *to*-infinitives (Rissanen 2000: 217f.).

A formal niche that remains unoccupied until the end of the EModE period is the progressive passive. First examples are not found before the late 18th century (see for example Rissanen 2000: 218; Denison 1998: 150ff.; Kranich 2010: 116ff.; Smith and Rayson 2007: 130; Hundt 2004b: 92ff.). Rissanen states that after it had appeared, "the set of progressive forms in all tenses, active and passive, is fully developed" (2000: 216).

Analysing the progressive's frequency development in different linguistic contexts (present, past, perfect, modal, modal perfect and non-finite forms), Kranich observes that once

[16] See also Petré (2017) for a corpus study of EModE progressive uses in present tense main clauses.

progressive use in these contexts was established, the relative frequencies remain relatively stable (2010: 126, Table 8). This leads her to conclude that

> the kind of change that might have been expected in a grammaticalization process, i.e. that the use of more unusual combinations, e.g. with modal and/or perfect, would become more common, is not evidenced in the data. [...] The use of the progressive in the present and past tense remains predominant, with present tense occurrences even becoming slightly more common in the 20th-century data. [...] On this basis, I would like to suggest that grammaticalization does not necessarily lead to frequency changes in the paradigmatic distribution, only to an extension in the possibilities of paradigmatic use (as in the new progressive passive [...]); otherwise, a construction will simply grow in frequency especially in those contexts in which its function is particularly often required (2010: 127f.).

Similar observations are made by Smitterberg (2005: 139, Table 48), who – despite the progressive's overall frequency increase in the 19th century – does not detect any clear signs of progressive verb phrases becoming formally more complex (except for the progressive passive). He states that "in fact, the percentage of all progressives [...] that incorporate no modal, perfect, or passive auxiliaries increases over the century" (2005: 139f.).

Thus, from a formal side, the most noteworthy development in the use of the progressive in ModE is certainly its combination with the passive, which Denison regards as "one of the few clearcut grammatical innovations of lModE" (1998: 152). The first unambiguous – and frequently cited – example goes back to 1772:

(3.2) *I have received the speech and the address of the Hous of Lords; probably, that of the of the House of Commons* ***was being debated*** *when the post went out.* (J.Harris, in Ser. Letters 1st Early Malmesbury, Vol. I, p.264, Denison 1998: 152)

The use of the progressive passive evoked strong prescriptive resistance in the 18th and 19th centuries. It was considered unnecessary since an alternative progressive construction that expressed passive meaning, the co-called *passival* (3.3), was well-established at the time. (Smith and Rayson 2007: 130; Hundt 2004b: 92).

(3.3) *The house* ***is building.*** (Hundt 2004b: 79)

As Rissanen makes clear, there was "little risk of confusion between the active and the passive meaning [...], as the subject is normally antimate in the former case and inantimate in the latter" (2000: 218).

While the progressive passive shows a major frequency increase from the 19th century onwards, the use of the passival has continuously declined (Hundt 2004b: 104, Figure 8). However, as Hundt (2004b) shows, it was not the progressive passive that ousted the passival. The latter was already in decline before the progressive passive was first used. Instead, the growing acceptance of present progressives to occur with inantimate

subjects (Hundt 2004a: 60, Figure 4a) resulted in major restrictions regarding the use of the passival: Its use with inantimate subjects became potentially ambiguous, which led to its consequent decline (Hundt 2004b: 113).

Smitterberg's CONCE data shows that the passival was already infrequent at the beginning of the 19th century and that it basically went out of use over the following one hundred years. His numbers for the progressive passive indicate that the construction established itself slowly but steadily in the course of the 19th century (2005: 128, Table 41).

After having summarised the supposedly most central of the progressive's developments – its grammaticalization as aspect marker and the development of its formal paradigm – the next section will take a look at a development that is less frequently addressed but no less important: the development of the progressive's non-aspectual uses.

Development of non-aspectual functions

The progressive cannot only be used as a marker of aspect but also to express a speaker's subjective viewpoint[17] and with future time reference (cf. chapter 3.4.2 & 3.4.3).

I already made reference to Killie (2008), who reports that OE and ME progressive use is characterised by much higher relative frequencies of non-aspectual subjective progressives than EModE.

Kranich (2010: 2002ff.) distinguishes between three special progressive functions. First, the so-called *ALWAYS*-type, where the progressive combines with an adverbial like *always* and frequently expresses a negative evaluation of the situation (*He is always/constantly complaining*). Second, special progressives without an adverbial of the *ALWAYS*-type, which generally express emphatic meaning (*You're driving me crazy!*). And finally, so-called interpretative progressives, functioning as an interpretation of the situation (*When he placed the flowers on your desk, he was apologising*).

In the ARCHER-2 data, Kranich finds that the overall share of these uses among all progressives amounts to 33% in the first half of the 17th century, while the aspectual functions account for 67% of all progressives. By the end of the 18th century, the share of special progressives has decreased to 6% (with 94% aspectual functions) (2010: 228, Table 39). Kranich concludes that this decrease in the use of special progressives must be linked to the construction's increasing objectification/de-subjectification in the course of its grammaticalization as aspect marker (2010: 241,245; 2013: 27).

Interestingly, though, from the 19th century onwards, the share of special functions has constantly been on the increase again (accounting for 14% of all progressives in the second half of the 20th century) (2010: 228, Table 39). Thus, in the course of the past 200 years, progressive usage has become more subjective again, while still not attaining the level exhibited 400 years ago. According to Kranich, this development is largely due to the emergence and increasing use of the interpretative function, which has gained ground in the 19th and 20th centuries, and to the fact that the *ALWAYS*-type progressive has become increasingly associated with subjective meaning (2010: 247,252).

[17] I refer to these uses as "special uses".

The special progressive uses without an *ALWAYS*-type adverbial are the ones that seem to be most similar to emphatic, non-aspectual progressives in OE and ME. As Kranich points out, these uses seem "to linger throughout the history of the progressive without changing much" (2013: 26). However, in spite of the recent resurgence of subjective progressive usage, special progressives without *always* or a related adverb have continued to decrease in relative frequency (2010: 219, Table 34).

Thus, it can be summarised that the period up to 1800 – in which aspectual progressive use increased in frequency due to the progressive's grammaticalization as aspect marker – was marked by a de-subjectification of progressive use. In this respect, Kranich explains:

> We can thus say that the overall process of secondary grammaticalization is not accompanied by subjectification, as has been claimed repeatedly (see e.g., Wright 1994), but rather by objectification or de-subjectification, i.e., the loss of possible contexts for speaker-based meanings (2013: 27).

However, after the progressive's grammaticalization was complete, subjective uses became more frequent again (while the total number of aspectual uses is also increasing). Kranich argues that this re-subjectification is not connected to the progressive's grammaticalization. It follows the grammaticalization process rather than being part of it (2010: 243ff.).

Smitterberg (2005: 207ff.) analyses the development of special progressive uses during the 19th century. Similar to Kranich, he distinguishes three functions (*ALWAYS*-type, interpretative function and so-called potentially experiential progressives[18]). In the CONCE data, only the potentially experiential progressives exhibit a statistically significant increase (from 10% of all progressives to 17%) (2005: 222, Table 74). The interpretative progressives (non-significantly) increase in overall and relative frequency (from 3% of all progressives to 6%) (2005: 231, Table 77). Thus, Smitterberg's findings neither confirm nor contradict Kranich, who reports increasing use of interpretative progressives.[19]

Apart from subjective meaning, non-aspectual progressives can also express future time reference.[20] Present progressives can refer to future plans and arrangements (*I'm meeting Richard on Sunday*) and are frequently referred to as *futurate* progressives. Second, combinations with *will* or *shall* (*I'll be returning home soon*) frequently do

[18] The experiential functions are similar to Kranich's subjective functions without *ALWAYS*-type adverbial. The term *experiential* is also used by Wright (1994 & 1995).

[19] However, Smitterberg's findings might be flawed by his methodology: He identifies two of the three subjective functions (the *ALWAYS*-type and the potentially experiential progressives) solely on the basis of formal criteria (2005: 213,217). Such a formal approach to progressive functionality is not without problems: For example, it seems unlikely that all progressives that combine with an adverbial of the *ALWAYS*-type actually express subjective meaning. Consequently, it may well be the case that a substantial number of Smitterberg's *ALWAYS*-type progressives are actually not subjective in meaning at all. This concern is voiced by Kranich, who tests Smitterberg's formal criteria and concludes "that […] [they] are not reliable indications" (2010: 211). Thus, Smitterberg's results for the *ALWAYS*-type and potentially experiential progressives should be treated with caution.

[20] It should be noted, though, that the progressive's future uses can be derived from aspectual usage (Kranich 2013: 19, Figure 1). See also chapter 3.4.3.

not express aspectual meaning but present the future situation in its entirety and as occurring as a matter of course (Leech 2004: 67).

Nesselhauf uses ARCHER to study the development of futurate progressives and finds that "if both the clear and possible instances of future time reference are taken into account, the progressive futurate has roughly tripled in frequency between 1750 and 1990" (2007: 196). Regarding genre-specific usage, she shows that "the most substantial increase has taken place in the speech-based registers of drama and fiction conversation. [...] In the more formal written text types [...], practically no increase can be observed" (2007: 198). Since the share of futurate progressives among all progressive tokens is comparatively low, Nesselhauf concludes that "the 'future' use of the progressive cannot be claimed to be the decisive factor in the development of the construction" (ibid.).

Kranich uses the term "derived aspectual functions" to refer to progressives with future time reference. She distinguishes between futurate progressives as expression of the near future and *will/shall* progressives as "matter-of-course" futures (2010: 179ff.). Her data suggests that the progressive as an expression of near future "has been on the rise since the second half of the 19th century" (2010: 186). She argues "that the increasing use of the progressive with near future reference has had its share in the general rise of the construction in recent times" (ibid.). Regarding the matter-of-course future, Kranich's results indicate that "uses of *will/shall* + progressive, while being overall quite infrequent [...], show a rise in the last half-century" (2010: 186). The fact that both future uses have reached a certain frequency level only in recent times leads Kranich to conclude that their grammaticalization must have happened not long ago (2013: 23).

Celle and Smith (2010) study the development of *will/shall* + progressive. They show that the combination has been available since at least the ME period. However, the non-aspectual meaning is not found until the 19th century (2010: 257) and does only develop a foothold in British and American English in the second half of the 20th century (2010: 242ff.). Thus, the grammaticalization and frequency increase of the construction is indeed a very recent (and probably still ongoing) phenomenon.

After having reviewed the progressive's long-term diachronic development, I will now move on to its more recent development in the 20th century, which constitutes the immediate background of the present study.

3.3. 20th-century developments

The progressive's development in 20th-century English has received considerable scholarly attention (e.g. Mair and Hundt 1995; Smith 2005; Mair 2006b, chapter 4.3; Leech et al. 2009, chapter 6). It is characterised by a continuation of the construction's long-term frequency increase but also by specific functional and formal developments.

3.3.1. Frequency development

Written English

Hundt's ARCHER data reveals a dramatic increase of progressive frequencies in BrE between 1900-49 and 1950-90 from M=173 to M=330 (2004a: 58,69). For AmE the 1900-49 data point is missing, but between 1850-99 and 1950-90 an increase from M=192 to M=301 can be observed. Thus, progressive frequencies have increased in both national varieties in the course of the 20th century, with a slightly higher level of use in BrE in the second half of the century.

Kranich's ARCHER-2 figures for 20th-century BrE usage (2010: 95), presented in Table 3.1, differ considerably from those of Hundt. However, they also indicate a rapid frequency increase: from M=350 for 1900-49 to M=393 for 1950-99. In the second half of the 20th century, the progressive reaches highest levels of use in the speech-based genre of Drama, followed by Fiction, which is characterised by repeated use of dialogue, and the involved and supposedly informal genre of private Letters. Also News and – somewhat surprisingly – the conservative genre Religious Sermons attain high M-coefficients. The findings for the latter genre are, however, based on a very small sample and likely to be skewed by idiosyncratic preferences (2010: 101, Table 5).

The first study explicitly designed to track short-term and ongoing changes in progressive usage is Mair and Hundt (1995). Analysing progressive frequencies in the Press sections of the Brown family corpora[21], the authors reveal a statistically significant increase in BrE (606 tokens in LOB vs. 716 tokens in F-LOB). Likewise, in AmE the progressive significantly increased from 593 tokens in Brown to 663 tokens in Frown. The differences between BrE and AmE fail to reach statistical significance.

In a later paper (Hundt and Mair 1999), the authors emphasise that the progressive's recent frequency development is by no means steady across all domains of language use, but – very much like its long-term development – has to be seen as mediated through genre. Unlike in the innovative Press sections of the Brown corpora, progressive frequencies remain stable on a low level in the conservative genre of Academic Prose (1999: 229f.).

Analysing all genres of the Brown corpus family, Smith (2002 & 2005) provides a very detailed account of 20th-century progressive usage. Like Mair and Hundt, he reports an overall frequency increase both for BrE and AmE. His results are summarised in Table 3.2, which reports frequencies per million words (pmw).

As in Mair and Hundt's findings (1995), slightly higher levels of progressive use can be observed for BrE. The overall rate of change, however, is basically equal for the two varieties. Regarding the synchronic and diachronic variation in BrE, Smith points out two tendencies:

> (a) a diachronic tendency of increasing use of the progressive across the *non-fiction* (or "informative") categories of the corpora; and (b) a synchronic

[21] Four matching one million-word reference corpora for BrE (LOB (data from 1961) and F-LOB (data from 1991)) and AmE (Brown (data from 1961) and Frown (data from 1992)).

Variety	Genre	1961	1991/1992	Change
BrE (LOB & F-LOB)	Press	3,244	3,817	**+17.7%
	Gen prose	2,158	2,534	**+17.4%
	Learned	844	952	+12.8%
	Fiction	5,199	5,146	–1.0%
	Overall	2,916	3,176	**+8.9%
AmE (Brown & Frown)	Press	3,131	3,464	+10.6%
	Gen prose	2,018	2,181	+8.1%
	Learned	1,247	896	**–28.2%
	Fiction	4,727	5,395	**+14.1%
	Overall	2,782	3,024	**+8.7%

Table 3.2.: Progressives pmw in the Brown family corpora (adapted from Table 4 & 13 in Smith 2005: 67,79).

> tendency for the progressive to be most common in *fiction* and next highest in the press categories of reportage and editorials. (2005: 67)

With regard to the AmE data he observes:

> the pattern across registers is rather more mixed than in British English. Indeed, learned writing shows a substantial decrease, in contrast to the stable frequency of British English. However, synchronically, the rank ordering of the broad genre groupings is the same as in BrE, that is, learned writing is the lowest-scoring category, fiction the highest, and press the second highest. (2005: 79f.)

Smith's findings are in accordance with those of Biber et al. (1999: 462, Figure 6.4), who report highest levels of progressive use for Fiction, followed by News and Academic Prose.

Smith regards the repeated use of dialogue in Fiction as a likely explanation for the high progressive frequencies, since it is well known that the construction is most readily used in spoken conversation. Thus, "genres with the most conversation-like content will favour progressives" (2005: 68). A second reason is sections of narrative prose, in which the progressive is frequently used in background descriptions such as (3.4):

(3.4) *She **was pouring** a sherry for Aunt Florence when the door swung open and Carola made one of her grand entrances.* (Smith 2005: 68)

The low frequency of the progressive in Learned/Academic writing can be linked to situation types. As Smith points out "the situations referred to by writers in this genre are more typically characterised as having general validity, rather than as ongoing at a particular point or interval in time. For the former, the simple present will usually be preferred over a present progressive" (2005: 68).

With regard to the diachronically stable use of the progressive in BrE Fiction, Smith hypothesises that the genre had already incorporated new functional developments of the progressive in the early 1960s, while this process might still be under way in other genres (2005: 69). If this was the case, though, the development in AmE would seem to be lagging behind, since Fiction is the only genre in this variety that exhibits a significant increase in progressive frequencies. However, given the generally very similar levels of progressive use in BrE and AmE genres, this explanation does not seem very compelling.

Drawing on supplementary corpus material[22], Smith presents figures for progressive frequencies in Social Letters (i.e. letters between intimates) and makes the striking observation that progressive numbers in this informal written genre are higher than in any other written genre and also than in all but one spoken genre (Telephone Conversation) (2005: 73). The construction almost reaches the level of 10,000 tokens pmw, which is an extraordinarily high number. According to Smith there are two factors that appear to promote the use of the progressive in this genre: "a frequent concern with relaying news of current, ongoing events (including planned, imminent events) [...] and a focus on current emotional states [...]" (2005: 75). Yet, one has to bear in mind that the figures are based on a small data sample and need not necessarily be representative.

Leech, Hundt, Mair and Smith (2009) corroborate the overall and genre-specific results of Smith (2005) for written English, concluding that "[b]oth national varieties present a picture of the progressive relentlessly marching on, much as it has been over the last centuries" (2009: 122).

Thus, the progressive's recent frequency development in written English has been thoroughly researched, and there can be no doubt regarding the construction's continued increase. Let us now turn to the spoken modality, in which non-standard changes are generally assumed to originate and to be more advanced than in the written medium.

Spoken English

Results concerning the progressive's recent frequency development in spoken English are based on much less data than the findings for written English. However, given the fact that corpora containing faithful transcriptions of spoken language have not been available until fairly recently, this situation does not come as a surprise.

Smith (2005: 76ff.) analyses progressive use in two matching mini-corpora of spoken BrE, which were designed to mirror the diachronic relationship of LOB and F-LOB.[23] His results are shown in Table 3.3.

The data make clear that the progressive is used much more frequently in spoken genres than in written ones. While the average lies around 3,000 tokens pmw in written data, the average frequency in Smith's spoken data lies between 7,000 and 8,000 tokens

[22] Four files comprising 20,000 words from the Survey of English Usage corpus (SEU), which contain data from 1959-65, and eleven files comprising 22,000 words from the British component of the International Corpus of English (ICE-GB), which contain data from the early 1990s (Smith 2005: 73).

[23] The SEU-mini, an 80,000-word sample taken from the SEU corpus; and the ICE-GB-mini, an 80,000-word sample selected from ICE-GB (Smith 2005: 49f.).

	SEU-mini		ICE-GB-mini		Change in %
Genre	Raw freq	pmw	Raw freq	pmw	
Conversation	301	8,600	282	8,294	–3.6%
Broadcast Discussion	162	6,480	202	7,769	+19.9%
Live Commentary	103	5,150	171	8,550	**+66.0%
sports comm.	85	5,667	94	7,833	*+38.2%
ceremonial comm.	18	3,600	77	9,625	**+167.4%
Overall	566	7,075	655	8,188	*+15.7%

Table 3.3.: Progressives in two mini-corpora of spoken BrE (Table 10 in Smith 2005: 76).

pmw. As Smith points out, this "preponderance of the progressive in spoken language is echoed across the ICE-GB corpus as a whole [...] and supported by other studies" (2005: 76). He cites Biber et al. (1999: 462) and Mindt (2000: 248), who report a frequency of roughly 7,000 and 5,000 tokens pmw for spoken conversation respectively.

In Smith's mini-corpora the overall progressive frequency has significantly increased by 15.7%, with the highest increase in the Live Commentary genre. In Conversation, however, the use remains static, and also the conversational genre Broadcast Discussion does not exhibit a significant frequency increase. Smith hypothesises that

> It is possible, then, that by the early 1960s expansion of the progressive in conversation had already reached a plateau. [...] [T]he non-conversational registers of SEU-mini and ICE-GB-mini appear to be catching up with the high level of use attained by the informal registers of direct and telephone conversations [...]. (2005: 76ff.)

Smith also produces figures for the distribution of the progressive in the full ICE-GB corpus, which reveal that in Telephone Conversation the progressive is used even more frequently than in direct Face-to-face Conversation (ca. 12,000 tokens pmw vs. ca. 9,000 tokens pmw) (the same results are presented in Leech et al. 2009: 125, Figure 6.4).

Like Smith, also Leech et al. (2009: 124ff.) investigate the progressive's genre-specific frequency development in spoken 20th-century BrE. Using an adapted version of the DCPSE[24] (i.e. the DSEU and the DICE[25]), they compare progressive use in spoken BrE of the 1960s (DSEU) with the 1990s (DICE). Their results are summarised in Table 3.4.

While Smith observed an overall frequency increase of 15.7% in spoken BrE, Leech et al. report an increase of 38.5%. Furthermore, progressive use has increased significantly in four of the five genres, including Face-to-face Conversation. Thus, the authors

[24] The *Diachronic Corpus of Present-Day Spoken English*.

[25] Two mini-corpora of approximately 130,000 words each; designed to match LOB and F-LOB as closely as possible (Leech et al. 2009: 44).

	DSEU		DICE		Change in %
Genre	Raw freq	pmw	Raw freq	pmw	
Conversation (face-to-face)	437	6,293	576	8,906	***+41.5%
Telephone Conversation	68	6,890	123	12,935	***+87.7%
Broadcast Discussion	215	5,779	284	8,651	***+49.7%
Broadcast Interviews	37	6,917	35	5,892	–14.8%
Sports Commentaries	67	4,623	123	7,619	***+64.8%
Overall	824	6,043	1,141	8,368	***+38.5%

Table 3.4.: Progressives in two mini-corpora of spoken BrE (Table A.6.4 in Leech et al. 2009: 289).

conclude: "The still-changing frequency of the progressive in speech, especially in relatively unmonitored genres such as face-to-face conversation, may be a symptom of the construction continuing to grammaticalize, through an increase in its range of functions" (2009: 126).

The common denominator of Smith's and Leech et al.'s findings clearly is the progressive's much more frequent use in speech than in writing. Furthermore, the construction has become considerably more frequent in spoken BrE in the course of the second half of the 20th century, and this frequency increase has happened at a more rapid pace than in written English. Concerning the distribution and development in the different genres, however, the findings are contradictory. What is needed is a larger corpus on the basis of which the progressive's use and development in spoken BrE can be analysed more reliably.

The DCPSE (Aarts and Wallis 2006) constitutes such a corpus.[26] Using a revised, i.e. non-official, version of it[27], Aarts et al. (2010 & 2013: 17ff.) quantify the overall increase of the progressive in late 20th-century spoken BrE. Unfortunately, they do not focus on genre-specific developments but only on the progressive's overall diachronic change. Depending on the frequency metric, they attest a significant increase of 20.95% to 22.13% between LLC and ICE-GB.[28] Table 3.5 presents their findings pmw.

[26] It consists of two subcorpora: the LLC, which contains data from 1958-77 and comprises roughly 400,000 words, and the spoken subpart of the ICE-GB, which contains data from 1990-92 and also comprises roughly 400,000 words.

[27] Aarts et al. (2010: 158, footnote 10).

[28] Similar to Smitterberg (2005: 62), Aarts et al. (2013: 17ff.) advocate a more refined frequency metric than the number of progressives per hundred thousand words (i.e. the M-coefficient) or per million words. Instead they argue that progressive tokens should be measured relative to the number of contexts that offer an opportunity of use. Even though their reasoning seems convincing, it should be noted that their results for the progressive's frequency development remain more or less the same irrespective of the different methods of calculation. In order to ensure comparability to other studies, I decided to report Aarts et al.' findings pmw (see chapter 4.4.1 for a discussion of different frequency metrics).

	prog verb phrases	prog pmw	Change in %
LLC	2,973	6,406	
ICE-GB	3,294	7,824	*+22.13%
Overall	6,267	7,081	

Table 3.5.: The progressive's frequency development in DCPSE (adapted from Table 2.1 in Aarts et al. 2013: 18).

Even though the overall progressive frequency in ICE-GB is slightly lower than the values reported by Smith (2005: 76) and Leech et al. (2009: 289), the frequency increase reported by Aarts et al. lies roughly in between the values reported in the latter studies. The result corroborates the observation that the progressive has increased much more rapidly in spoken than in written language in late 20th-century BrE. In order to obtain a better understanding of this development, it will be interesting to see how progressive frequencies have developed in the individual genres of DCPSE (see chapter 5.1).

3.3.2. Explaining the frequency increase

Different explanations have been suggested for the progressive's recent frequency increase. The following sections focus on how exactly the progressive has changed in recent times and attempt to single out the most important driving forces behind the construction's increase.

The progressive's morphosyntactic paradigm

Smith (2005: 69ff.) and Leech et al. (2009: 124,288, Table A6.1) show that the progressive's frequency increase in written English has mainly occurred in connection with the present tense. While in BrE both the present progressive active (***+29.4%) and the present progressive passive (**+45.9%) have increased significantly, only the present progressive active has gained ground in AmE (***+33.5%), while the present progressive passive has decreased in use (**–33.0%). For spoken BrE Smith (2005: 77, Table 11) reports a significant increase of the present progressive active by 22.8%. Leech et al. report that

> we do know that the present progressive [active; *U.R.*] [...] is the realization *par excellence* in speech, and is becoming increasingly prevalent. It accounts for 85% and 90% of progressives in DSEU and DICE respectively. Since the corresponding figures in LOB and F-LOB are 36% and 43%, we may speculate that written BrE is increasingly shifting towards spoken norms of tense and aspect combination. (2009: 126f.)

The present progressive's[29] increase is significant in all genre categories of LOB and

[29] From now on, if not stated otherwise, the term "present progressive" refers to the present progressive active only.

F-LOB (Smith 2005: 84). In Smith's spoken data the increase also occurs in all genres. However, it is only significant in Live Commentary (2005: 77, Table 12).

Regarding the relationship of the present progressive and its competitor construction, the simple present, Smith observes that "[b]oth constructions are increasing in frequency, although the present progressive is increasing at a faster rate" (2005: 85). This means that speakers increasingly choose to use a present progressive where they might have chosen a simple present before.

Another observation concerns the fact that the present progressive frequently occurs in colloquial speech. As Leech et al. (2009: 128) point out, the proportion of text within quotations (most of it direct speech) has increased significantly in the Brown family corpora, which poses the question whether the spread of the present progressive in written language is simply an artefact of this development. However, the authors show that this is not the case but that "use of the progressive outside speech quotations has increased at a faster rate than within speech quotations" (2009: 128) (see also Smith 2005: 85ff.).

Analysing not only LOB and F-LOB but also BLOB[30] and BE06[31], Smith and Leech (2013: 88f.) show that the use of the present progressive remained stable until 1961 but has increased ever since. A particularly steep increase can be observed for Press writings between LOB and BE06. These findings are important since they show that the frequency differences between LOB and F-LOB are not just short-term fluctuations but part of a long-term trend.

Apart from the present progressive, changes have also been observed for the past progressive (active). While its use has remained stable in written AmE, it has significantly declined in BrE (*–8.8%) (Leech et al. 2009: 288, Table A6.1; Smith 2005: 70,80 Table 5&14). For spoken BrE Smith reports a non-significant increase (2005: 77, Table 11) while Leech et al. (2009: 127) report a significant one. Thus, while the past progressive has decreased in written BrE, no such trend can be observed in the spoken data. Due to the very limited size of the spoken corpora, these results should be regarded as preliminary, though. Attempting to interpret these figures, Leech et al. (2009: 127) refer to past progressives as the one in (3.5):

(3.5) *I **was talking** to this guy at college and uhm he's really really really boring.*
(Leech et al. 2009: 127)

They point out that "Past tense speech-reporting progressives such as [(3.5)] are typical of narrative style in face-to-face and phone conversations, for instance, but as far as we are aware they are not yet widespread in printed BrE" (2009: 127). Increasing use of such speech-reporting past progressives in spoken conversation might play a role in the alleged frequency increase of the past progressive in spoken BrE.

The last area for which significant distributional changes of the progressive across the morphosyntactic paradigm can be observed are combinations with modal auxiliaries. While most of the core modals have declined substantially in connection with the simple

[30] A prequel to LOB that contains data from 1931 (Leech and Smith 2005).
[31] A sequel to F-LOB that contains mid-2000s BrE texts (Baker 2009).

aspect (cf. Leech et al. 2009: 71ff.), this is not the case for combinations of modals and the progressive. Several publications have addressed this phenomenon (Leech et al. 2009: 139ff.; Smith 2003; Smith 2005: 70,139ff.; Celle and Smith 2010; Smith and Leech 2013: 86ff.) and report two important trends: First, while modal + progressive patterns have not become more frequent in AmE, their use has increased significantly in BrE (*+29.5%) (Leech et al. 2009: 288, Table A6.1). Second, most of the increase in BrE goes back to the pattern *will + be -ing*. *Will* is the only modal that has significantly increased in connection with the progressive between LOB and F-LOB (*+35.0%) (Leech et al. 2009: 295, Table A6.14). As Smith (2003) and Celle and Smith (2010) show, this increase is a recent development (cf. chapter 3.2.3) and linked to a functional change. Despite their aspectual marking, the majority of cases of the pattern *will + be -ing* do not convey the meaning "event in progress at a future time". Most of the uses are aspectually underspecified and present the future event as happening as a matter of course.[32]

Smith and Leech (2013: 87) show that modal + progressive combinations have not only increased between LOB and F-LOB but over a time span of 75 years. While modal + progressive combinations occur with a frequency of roughly 100 tokens pmw in BLOB, the number is way above 200 tokens pmw in the BE06 corpus. Thus, similar to the development of the present progressive, we are witnessing a long-term trend.

As for the progressive passive, no such long-term trend can be observed. While the present progressive passive significantly increased in use between LOB and F-LOB (Smith 2005: 122ff.; Smith and Rayson 2007: 136ff.; Leech et al. 2009: 136ff.), this trend is not corroborated by findings of Smith and Leech (2013: 85f.), who show that the overall level of use is basically identical in their 1931 BLOB and their 2006 BE06 data. The increase between LOB and F-LOB seems to have been a mere fluctuation.

Regarding the progressive passive's use in AmE, Smith (2005: 80, Table 14) and Leech et al. (2009: 288, Table A6.1) report a significant decline of about 30% for the present progressive passive. In the light of this observation Leech et al. note:

> It is also noticeable that the *non*-progressive passive has declined significantly further in AmE than in BrE [...]. This leads us to speculate that the diminishing use of progressive passives in that variety may be the result of a stronger prescriptive resistance in the United States to use of the passive in general. (2009: 136)

A very recent phenomenon is the progressive *get*-passive as in (3.6).

(3.6) *... but my point is that all Waterloo pupils* ***are getting blamed****.* (Smith and Rayson 2007: 140)

As Hundt states, "[t]he first unambiguous *get*-passives in ARCHER are from the twentieth century" (Hundt 2004bb: 99). She shows that compared to the progressive *be*-passive the construction is a marginal pattern in BrE that does not seem to be

[32] See also chapter 3.4.3.

increasing in use. She finds four tokens in LOB and F-LOB respectively (and two in each of the two AmE corpora).

Soete (2014) studies the recent development of a phenomenon that has not received much attention so far: combinations of the so-called *going to*-future and a progressive infinitive (3.7).

(3.7) ***I'm going to be doing*** *an extra year at med school anyway.* (Soete 2014: 8)

Even though the overall frequencies of the pattern are low, Soete's COHA[33] results suggest two interesting conclusions: First, very much like the progressive *get*-passive, the pattern is a relatively recent addition to the progressive's morphosyntactic paradigm. First uses are attested for early 20th-century data. Second, unlike the progressive *get*-passive, the pattern's use has become more frequent in the course of the 20th century (Soete's numbers for contemporary AmE range between 5 and 15 tokens pmw, while they were basically zero at the beginning of the 20th century).

While such a pattern is certainly not contributing to the progressive's frequency increase in any substantial way, it shows that apart from quantitative changes the progressive's morphosyntactic paradigm is still prone to structural changes.

A similar point is made by Mair (2006b), who explains that

> the twentieth century has seen the creation of new progressive constructions in the few remaining niches of the verbal paradigm in which the form did not used to be current in the recent past – in the main forms such as the present or past perfect passive progressive ('*I have been being interviewed*'), the future/conditional/modal passive progressive ('*I will/would/might* etc. *be being interviewed*') or the future/conditional/modal perfect passive progressive ('*I will/might* etc. *have been being interviewed*') [...]. These new progressive forms are, of course, highly salient to observers and interesting for a theoretical analysis, but statistically they are insignificant and certainly do not account for the global increase observed in the discourse frequency of the progressive. (2006b: 89f.)

This means that new progressive forms occur but do not contribute significantly to the construction's recent frequency increase, which is mainly driven by the most frequent form, the present progressive.

Functional developments and innovations

In chapter 3.2.3 I already mentioned Kranich's (2010: 202ff.) findings that indicate a re-subjectification of progressive use from the 19th century onwards. Regarding the 20th century, Kranich reports an increase in the so-called subjective, non-aspectual progressives from 11% to 14% (relative to the number of progressives) (2010: 228, Table 39). Interestingly though, of the three subjective uses distinguished by Kranich – hyperbolic *ALWAYS*-type, emphatic/attitudinal use without ALWAYS and interpretative use –

[33] *Corpus of Historical American English.*

only the interpretative use has substantially increased its share (from 29 tokens in the second half of the 19th century to 90 tokens in the second half of the 20th century) (2010: 223, Table 36). Kranich goes as far as claiming: "it can be gathered that the increasing use of the interpretative progressive is the main factor in the 20th-century rise of the progressive" (2010: 223). This conclusion, however, seems to overstate the statistical impact of the development, as a look at Kranich's overall progressive numbers makes clear. While the interpretative use certainly seems to be one contributing factor to the observed overall increase, it is nonetheless unlikely to be the major driving force behind it. Judging by Kranich's absolute numbers, it accounts for roughly 20% of the increase between the late 19th and 20th centuries.

Smith (2005: 105ff,165ff.) provides an in-depth analysis of so-called special uses of the progressive. These uses include futurate progressives and progressives that convey expressive or pragmatic meaning. These expressive and pragmatic progressives roughly correspond to Kranich's label of subjective progressives. The relative increase of the expressive functions is not very pronounced. While 8.6% of all present progressives in LOB belong to this group, it is 9.9% in F-LOB (2005: 106). According to Smith, a much higher level of use should be expected for spoken data since "spoken (especially conversational) discourse, by virtue of its interactive character, is likely to show a much higher rate of expressive usage" (ibid.).

Smith distinguishes between four different expressive/pragmatic uses: Resolute uses (conveying an emphatic intent by the speaker), attenuating/hedging uses (conveying a politeness/downtoning function), *ALWAYS*-type progressives, and interpretative uses. Since numbers for the first two uses are very low, no sound conclusions regarding their development can be drawn. *ALWAYS*-type progressives increase from 12 tokens in LOB to 19 tokens in F-LOB. This might be indicative of an increase but could just as well be due to chance. In any case, the numbers are very low and, consequently, the *ALWAYS*-type progressive cannot be seen as a major contributor to the progressive's overall frequency increase. The only function that has undergone a clear frequency increase is the interpretative one. Smith identifies 52 clear cases in LOB, opposed to 97 cases in F-LOB. However, compared to the overall number of present progressives, the use of the interpretative function is still infrequent and its contribution to the overall increase is a noticeable but modest one.

Regarding the interpretative function's distribution across genres, Smith observes "that it is most common in fictional dialogue, which suggests that it is characteristically a speech-based feature" (2005: 196). Importantly, the increase still holds after removing all cases that occur inside quotations, which "strongly suggests that the colloquial feature has permeated the prose genres of the corpora, to the extent of occurring within learned texts" (2005: 197).

Smith regards the increase of the interpretative function not only as a sign of ongoing colloquialization of written norms, but also as a possible sign of grammaticalization. He states: "The increasing use of the interpretive progressive may be the result of a motivated extension of the basic aspectual function of the progressive, that of construing a situation as being 'in progress'" (2005: 198). Referring to Wright (1995), he invokes Traugott's concept of subjectification (1989 & 1995) according to which grammatical

meanings become increasingly based in the speaker's attitude towards what is said.[34]

Thus, both Kranich's and Smith's findings – far from being conclusive – seem to hint towards an ongoing (re-)subjectification of progressive use in 20th-century English, mainly based on the increasing use of the interpretative function.

Smith's findings for the *ALWAYS*-type progressive and the interpretative uses in LOB and F-LOB are replicated by Leech et al. (2009: 134ff.,296, Table A6.16). With respect to the increasing use of the interpretative function they assume that "it seems premature to conclude that interpretive progressives are a key factor in the spread of the present progressive".

Mair (2006b) points out a major methodological caveat regarding the identification of interpretative uses:

> In fact, attempts at statistical analyses in this case might even be self-defeating, because in a situation in which for the majority of cases the "interpretive" reading is one option alongside others only collecting the clear instances means under-reporting the phenomenon, whereas including all possible instances in the counts will lead to clear over-reporting. (2006b: 94f.)

This problem certainly has to be kept in mind with respect to Kranich's and Smith's findings.

Other progressive uses that are often mentioned in the context of the construction's recent frequency increase are progressives with future time reference. In chapter 3.2.3 it was noted that Kranich (2010) assumes these uses to be a contributing factor in the progressive's increasing use. In ARCHER-2 she finds 42 present progressives referring to the near future (futurate progressives) for the early 20th-century data and 56 tokens for the late 20th-century data. *Will/shall* progressives (matter-of-course futures) occur only 5 times in the early 20th century but increase to 22 tokens in the second half of the century.

Mair and Hundt (1995: 249, Table 1) report a non-significant increase of futurate progressives in the Press sections of LOB (25 tokens) and F-LOB (34 tokens). Frequencies are more or less stable for the AmE corpora (Brown = 11, Frown = 14).

Smith (2005: 113, Table 24), on the other hand, reports a slight downward trend from 56 futurate progressives in LOB to 45 in F-LOB. The function is most common in Fiction and nearly all tokens of this genre occur in speech quotations. According to Smith this "suggests that futurate use is a feature of colloquial or at least spoken language" (2005: 113). Tellingly, he reports substantially higher normalised frequencies for futurate uses in spoken BrE. What is more, in spoken BrE their use seems to be increasing diachronically (2005: 117, Table 27). It has to be noted, however, that the overall numbers are low and should not be overestimated (17 tokens for Conversation in LOB and 34 tokens in F-LOB).

Leech et al. confirm Smith's findings for written BrE and also provide figures for the AmE corpora. Based on a one-in-four sample they find 7 futurate progressives in Brown

[34] Note that this is slightly at odds with Kranich's (2010) view presented in chapter 3.2.3 according to which grammaticalization and subjectification are two distinct processes.

and 18 in Frown (2009: 294, Table A6.12). However, they regard it as "unlikely that the futurate use represents a source of continuing growth of the present progressive in BrE or AmE" (2009: 133).

Another function that has already been mentioned in connection with the progressive's long-term development (chapter 3.2.3) is the matter-of-course future, expressed by the pattern *will/shall* + progressive. The pattern itself has been available for centuries, the non-aspectual matter-of-course meaning[35], however, is a recent development. Celle and Smith (2010: 249, Table 4) find 15 non-aspectual uses in BLOB, 26 in LOB and 45 in F-LOB (very similar figures for LOB and F-LOB are provided in Smith 2003: 721, Table 5; Smith 2005: 152, Table 35; Leech et al. 2009: 296, Table A6.17). Similar to the interpretative progressive, the semantic classification of the pattern *will/shall* + progressive is a difficult and subjective endeavour, meaning that in many cases it cannot be decided whether the pattern expresses aspectual or non-aspectual meaning. Of those uses, however, that could be classified unambiguously by Celle and Smith, about 85% express the non-aspectual meaning. Even if all the unclear cases are taken into account, this meaning is still the dominant one. Thus, the matter-of-course meaning of the *will/shall* + progressive pattern clearly seems to be grammaticalizing.

The development of another specific functional pattern is addressed in Levin (2013). He analyses the *BE being adj.* construction (3.8), which indicates momentary agentive behaviour (cf. König 1995a). It is considered an interpersonal/subjective use of the progressive because it frequently makes reference to a speaker's evaluation of a situation. This is why several authors (e.g. Ljung 1980: 69ff.; Levin 2013: 193; Žegarac 1993: 214f.) regard the pattern as a special type of the interpretative progressive, by which the speaker interprets a behaviour/an action as momentary and deliberate (or insincere).

(3.8) *You**'re being naughty**.* (adapted from Levin 2013: 196)

So far, not much quantitative data addressing the pattern's development has been available. Leech et al. (2009: 292, Table A6.9) find 7 tokens of the present tense form of the pattern in LOB and 10 in F-LOB. For Brown they report 2 tokens, for Frown 14. Thus, on weak statistical grounds, one might suggest increasing use in written AmE.

Levin (2013) reports the development in AmE and his findings are based on considerably larger corpora. They corroborate the upward trend suggested by Leech et al.'s numbers. While in the 1920s subsection of the *Time* corpus, the pattern occurs with a frequency of only 1.1 tokens pmw, the frequency increases to 4.8 tokens pmw in the 2000s subsection. Even though the pattern remains an infrequent one, its frequency has increased in the course of the 20th century. Regarding the genre distribution, Levin finds highest numbers for Conversation (32 tokens pmw in the *Longman Spoken American Corpus*, LSAC), TV and radio shows (12 tokens pmw in *Corpus of Contemporary American English*, COCA) and Fiction (12.9 tokens pmw in COCA).

Pfaff et al. (2013) address a functional development that has attracted relatively little attention so far: The past progressive as an expression of the recent past, as in example (3.9).

[35] See Celle and Smith (2010: 248ff.) for a fine-grained functional analysis.

(3.9) *I **was just reading** the other day that Billy Graham, as you say he's been an important part of Nixon's life, he officiated at the funeral of Hannah Nixon, who died in 1967.* (Pfaff et al. 2013: 219)

According to the authors, such uses of the progressive are perfective in nature and incompatible with the aspectual meaning of the construction. The main function of such uses is to convey the notion of temporal recentness. The authors report a significant frequency increase in COCA (211 tokens in the early 1990s, opposed to 283 tokens in the early 2000s.) (2013: 224, Table 9.2). They conclude that the structure is currently gaining ground and argue that "the past progressive is increasingly used in contexts of recent past time and might even be developing into a fully fledged marker of recent past" (2013: 218).

In the light of the overall decline of the past progressive in written BrE this is an interesting hypothesis. It could be speculated that this overall decline is due to a narrowing meaning range of past progressives to situations of recent past. However, the hypothesis is rendered somewhat implausible by the fact that Pfaff et al. studied AmE, which – unlike BrE – does not exhibit an overall decline of the past progressive.

Summing up the results presented in this section, it can be stated that the progressive's functionality has changed to a certain but no dramatic degree in the course of the 20th century. Interpretative uses seem to have increased fairly substantially, but not to an extent that would make them the major driving force behind the increase of the present progressive. The so-called matter-of-course future is another area of use that exhibits a clear frequency increase. However, similar to the interpretative uses, the gains are not substantial enough to regard the function as a major factor in the recent spread of the progressive. Altogether, no single functional development can be identified that would be pronounced enough to account for the (present) progressive's overall frequency increase.

The progressive with stative verbs

Linked to changing progressive functionality are changing verbal preferences of the progressive. For example, *be* can be used agentively in the interpretative pattern *BE being adj.* as in example (3.8). Given the general assumption that stative verbs do not readily lend themselves to progressive interpretation, such uses might be indicative of the frequently heard claim that the 20th century has seen a greater readiness to use the progressive with exactly such verbs.

Increasing use with stative verbs could be taken as an indication that the construction is developing away from a marker of progressive aspect towards a general imperfective marker, which no longer denotes dynamicity (cf. Comrie 1976: 38f. & 1995: 1245; Bybee et al. 1994: 141; Kranich 2010: 32ff.).

Corpus data for 20th-century written BrE is provided by Smith (2005: 96, Table 23). Leech et al. provide data for BrE and AmE (2009: 292, Table A.69). Both studies analyse the use of the present progressive with verbs that lend themselves to stative interpretation. Leech et al.'s results are summarised in Table 3.6.

Verb class	LOB 1961	F-LOB 1991	Brown 1961	Frown 1991/2
perception/sensation	8	13	4	10
cognition/emotion/attitude	32	31	21	33
being/having	21	33	19	28
stance	22	25	18	23
Overall	83	102	62	94

Table 3.6.: The present progressive with verbs lending themselves to stative interpretation (absolute numbers) (adapted from Table A.6.9 in Leech et al. 2009: 292).

The results reveal an increase in the use of stative verbs with the present progressive both in BrE and AmE. The development is slightly more pronounced in the latter variety. However, as Smith points out, "the overall frequency of the same verbs in the simple present has also increased" (2005: 96). In the light of the moderate gains of stative verbs Leech et al. conclude that "it seems that in printed English, use of the progressive with stative verbs did not contribute substantially to the growing use of the construction between the 1960s and 1990s, in either regional variety" (2009: 130).

Kranich classifies the progressives found in ARCHER-2 according to situation types and reports a very moderate relative increase of progressives referring to states (from 8% in the second half of the 19th to 10% in the second half of the 20th century) (2010: 153, Table 16).

Levin (2013) provides a very detailed study on the use of the progressive in written AmE with so-called private verbs. These verbs comprise two categories: First, mental states (*believe, like, dislike, hate, intend, know, love, pity, want, wish*) and second, mental and emotional processes (*expect, feel, hope, think, wonder*) (2013: 199). Even though not all of his classifications are easily comprehensible – why, for example, is *expect* classified as a mental process? – his results are nonetheless interesting.

The proportion of progressives among the mental state verbs increases from 6.6% in the 1920s section of the *Time* corpus to 10.9% in the 2000s section (2013: 201, Figure 8.4). This reflects an increased readiness to use these verbs in connection with the progressive. Furthermore, Levin points out that "[t]he number of types taking the progressive in the three decades, four (1920s), four (1960s) and seven (2000s), suggests that the spread to new types has only recently gained momentum" (2013: 200). The synchronic picture clearly reveals that the readiness to use a mental state verb with the progressive is by far highest in spoken Conversation (25% in the LSAC corpus), followed by Fiction (13.6% in COCA) and TV/radio shows (13.4% in COCA) (2013: 202, Figure 8.5).

For the supposedly less stative mental process verbs Levin provides the frequencies pmw. While the level of use is just above 20 in the 1920s section of the *Time* corpus, it almost reaches a value of 90 in the 2000s data. Especially *hope* massively increases its

probability of use with the progressive (2013: 204, Figure 8.6 & 8.7).[36]

In the light of these findings and those of Leech et al. (2009) it can be assumed that in the course of the 20th century the readiness to use the progressive with stative verbs has increased in written AmE and seems to have contributed to the overall increase in the construction's frequency. This might be taken as a hint that the progressive is indeed developing towards a general imperfective marker. For BrE the data base is less sound, allowing no such conclusion.

Finally, I would like to draw attention to a point made by Mair, who states that

> it seems plausible to regard the use of the progressive with stative verbs as an instance of contextually/pragmatically licenced rule-breaking for specific rhetorical or expressive effect – an option which has been available ever since the present system of rules emerged in the eighteenth century. The reason that the phenomenon seems more in evidence today than in the past is simply that the type of informal context in which it happens is less likely to have been preserved in the past (2006b: 92).

Mair provides early 19th-century examples and concludes that "the 'stative' progressives are thus not the recent innovation they are considered to be by many commentators" (2006b: 94).

Thus, the trend reported for stative progressives in AmE does not reflect genuine linguistic innovation but rather a statistical change that applies to already existing patterns of use.

Colloquialization: Changing stylistic conventions

While functional developments such as the interpretative use may have played their part in the recent rise of the progressive in written English, it has been made clear that they cannot solely be accountable for it.

Mair links the progressive's frequency development to his concept of *colloquialization*, which proposes that written norms are moving closer to informal norms of spoken usage (e.g. 2006b: 183ff.). He explains:

> As a linguistic term, [colloquialization] covers a significant stylistic shift in twentieth-century English: away from a written norm which is elaborated to maximal distance from speech and towards a written norm that is closer to spoken usage, and away from a written norm which cultivates formality towards a norm which is tolerant of informality and even allows for anti-formality as a rhetorical strategy. (2006b: 187)

Evidence for colloquialization is abundant. The process affects the textual macrostructure in that more and more quotations are used in written texts. On the microstructural level, colloquialization affects the choice between formal and informal grammatical constructions or lexical items (2006b: 188f.). For example, contractions have

[36] See also Anderwald (2017) for a recent study of increasing use of *I'm loving it* in AmE.

increased massively both in written AmE and BrE in the second half of the 20th century (2006b: 189, Table 6.2). Also Leech, Smith and Rayson (2012) report several developments indicative of colloquialization. For instance, a substantial decline of the passive voice in BrE between 1931 and 1991. Likewise, pied-piping – which, like the passive voice, is considered a feature of a formal, written style – is much less frequent in written BrE texts of 1991 than in texts of 1931. Further evidence is presented by Leech et al. (2009: 239ff.) and by Smith (2005: 201ff.), who provides a very comprehensive analysis of the colloquialization hypothesis.

There is agreement, however, that colloquialization is not the only major stylistic phenomenon that has shaped 20th-century English. Competing with it are a number of developments affecting the English noun phrase (e.g. an increased frequency of nouns, noun sequences and *s*-genitives) that are summarised under the term *densification* (Smith 2005: 230, Table 70; Mair 2006b: 192; Leech et al. 2009: 245ff.; Leech, Smith and Rayson 2012: 78f.).

According to Mair, the competing pressures of colloquialization and densification reflect "the ideal implicit in present-day English writing style in all genres [...] to maximize information density, but to avoid additional stylistic ornament of formality in order to give the impression that the resulting texts remain easy to read and accessible" (2006b: 193).

As Hundt and Mair (1999) point out, different genres show different degrees of "susceptibility" to colloquialization. While there are "agile" genres such as the Press section in LOB and F-LOB that are open to innovation coming from speech, there are more conservative genres such as Academic Prose that are much more "uptight" with regard to the incorporation of informal spoken features. Like several other features characteristic of spoken interaction, the progressive's frequency has not increased in the latter genre. While colloquialization has influenced language use in journalistic writing quite substantially, academic publications have much more been shaped by the pressure of densification.

In the light of these insights into the changing stylistic preferences of written English, Smith (2005: 201) poses the question whether the progressive's frequency increase is of itself a remarkable phenomenon or just a reflection of a wholesale drift towards a more speech-like style. Based on his corpus evidence he concludes that this is an unlikely scenario since colloquialization is not an all-encompassing phenomenon but at best "a patchy one" (2005: 228). This means that while it is unlikely that the progressive's frequency increase is completely driven by grammatical changes (i.e. new progressive forms and meanings), it is also unlikely that colloquialization alone accounts for the construction's increasing use. Both factors play a role and are hard to disentangle. One example is progressive use with stative verbs: On the one hand, it might reflect a growing readiness to use the progressive in contexts in which it is not obligatory, indicating grammatical change. On the other hand, in written texts it can clearly be seen as a sign of colloquialization. This observation is reflected in Leech et al.'s statement that "colloquialization builds on grammaticalization" (2009: 244).

A change under its own momentum?

This subchapter has presented contributing factors to the progressive's recent frequency increase. It has been shown that the progressive can be used in new formal contexts and with a range of non-aspectual meanings and that some of these factors (for example the interpretative progressive and, in AmE, use with stative verbs) seem to contribute to the frequency development. However, it has also been shown that all these supposedly special uses are not frequent enough to fully account for the progressive's overall increase. The situation is summed up by Mair, who states that

> the increase in the textual frequency of the progressive has largely occurred within the existing framework of forms and rules. Both the new forms (e.g., the present perfect passive progressive) and the suspected new uses (progressive for stative verbs) are far too infrequent to account for the frequency shifts that can be observed. (2006b: 89)

Similarly, Leech et al. ask: "Why is the progressive [...] still increasing apace, including in those syntactic environments in which it has long been common?" (2009: 269). They, too, attribute a certain effect to new functional uses such as the interpretative progressive but, more importantly, formulate the following hypothesis:

> It is tempting to adopt the thesis that frequency changes take place under their own momentum: that increase begets increase, and decrease begets decrease, unless other factors interfere with the change. The argument, which sees interaction between use (communicative events) and user (cognitive resources) as a basis for language changes, runs as follows. Increasing frequency feeds into the expansion of the language system through making peripheral combinations less peripheral, or central ones more central. As a result, users' resort to those usages is still more frequent (2009: 269f.).

According to this view, the progressive's increased frequency is not only an *effect* but at the same time a *cause* for even more frequent use. In a kind of positive feedback loop the progressive is gaining strength under its momentum – a snowball effect in language change.[37]

While such a process definitely has its limits and is held in check by competing forces, it is certainly compatible with research in usage-based linguistics. Leech et al. (2009: 270, footnote 36) refer to the work of Bybee, which regards frequency as a major explanatory factor in language change. Very much in line with Leech et al.'s hypothesis is Bybee's insight that "repetition strengthens memory representations for linguistic forms and makes them more accessible" (2007: 10). Regarding the question of whether frequency is a cause or an effect Bybee explains: "On the one hand, frequency is just a tally, a

[37] This idea is not entirely new. As early as 1982, Strang mentions it as a possible explanatory factor for the progressive's frequency development in early 19th-century English: "These, I think, are the main ingredients of the increase, and the growth-period for all of them is in the year around 1800: *there is probably some consequential snowballing* [my emphasis; *U.R.*], and some growth of specialised idiomatic uses, which I have not investigated" (1982: 446).

pattern observable in texts, which is of course an effect. On the other hand, frequency or repetition of experiences has an impact on cognitive representations and in this way becomes a cause [...]." (2007: 18)

At this point, I would like to conclude the analysis of the progressive's diachronic development. The next section provides a systematic overview of the construction's uses in PDE and draws attention to lexical aspects of progressive use.

3.4. Functionality and use in present-day English

The following section aims to provide a systematisation of the progressive's present-day semantics (mainly based on Kranich's (2013) classification). Since most progressive functions have already been addressed in previous sections of this chapter, I will avoid lengthy discussions and present the different functions in what could be called a simple reference list.

3.4.1. Aspectual functions

Progressive aspectuality The default meaning of the progressive is the expression of progressive aspectuality, i.e. conveying the meaning that a dynamic situation is in progress at topic time (topic time being the time for which a particular utterance makes an assertion) (cf. Klein 1994: 37; Kranich 2013: 12; Huddleston and Pullum 2002: 162). The situation is presented from within (i.e. as imperfective) and can be characterised as dynamic and ongoing.[38] Since the situation is presented as ongoing, it is also presented as being not necessarily complete. A further strong implication resulting from ongoingness and incompletion is the notion of limited duration/temporariness (cf. Leech 2004: 18ff.; Quirk et al. 1985: 198). All this applies to the progressive in (3.10).

(3.10) *Sandra **was running** down the street when the accident happened.*

Imperfective aspectuality "The major difference between progressive aspectuality and imperfectivity in general is that the former is inconsistent with a purely static situation: it conveys some measure of dynamicity." According to this explanation by Huddleston and Pullum (2002: 167), imperfective situations need not necessarily be dynamic but can be static in nature. Thus, progressives referring to situations that hold at topic time but that are not properly dynamic can be said to express imperfective aspectuality and are often called general imperfective uses (cf. Kranich 2013: 13). The progressives in example (3.11) express general imperfective meaning but are incompatible with a "dynamic situation in progress" reading.

[38] This view, however, is not unanimously accepted. Croft, for example, claims that "[t]he Progressive construction [...] alters the aspectual contour to a (transitory) state." (2012: 152) (see also Michaelis 2004: 4). While it is certainly the case that the progressive can – with certain verbs – refer to transitory states, the present work argues for the generally accepted view that most progressive uses refer to dynamic, i.e. non-stative, situations.

(3.11) *I can see them from where I'**m sitting.** [...] He **is wearing** outdoor clothes and carries a large brown paper bag.* (Kranich 2013: 13)

Iteration/habit Progressive uses that refer to temporary habits are less prototypical but can nonetheless be regarded as aspectual uses (cf. Kranich 2013: 12). In the following example, the habit has limited duration and refers to a situation that is not yet complete.

(3.12) *I'**m taking** dancing lessons this winter.* (Leech 2004: 32)

Leech identifies another habitual use, which he defines as "repetitions of events of limited duration". He explains: "Here the notion of limited duartion applies not to the habits as a whole, but to the individual events of which the habit is composed" (2004: 33).

(3.13) *Whenever I pass that house the dog'**s barking.*** (Leech 2004: 33)

3.4.2. Special/non-aspectual uses

There are several uses of the progressive that do not (primarily) express an aspectual meaning. Some of these uses serve pragmatic functions by conveying the speaker's subjective viewpoint of the situation. Others are somewhat diffuse in meaning and hard to define properly.

***ALWAYS*-type progressive** If the progressive combines with an adverbial like *always*, *constantly* or *forever*, it frequently expresses a hyperbolic, negative evaluation of the situation (cf. Kranich 2013: 17; Mindt 2000: 249). Consider the following example in which the speaker expresses a negative attitude towards the behaviour of the referent. From a purely aspectual point of view, the simple aspect would be perfectly fine. Then, however, the pragmatic implication of annoyance would be lost.

(3.14) *You'**re always having** the best things.* (Mindt 2000: 249)

Interpretative progressive A concise definition of interpretative progressive meaning is given by Ljung (1980: 70f.): "The A part expresses the observed behaviour, the B part sums up or interprets this behaviour and the predicate used for this summing up or interpreting is invariably put in the progressive".[39] He provides the following example:

(3.15) *On p.21 Chomsky says: "[...]". This may seem innocuous enough, but what he **is really saying** is that [...].* (Ljung 1980: 71)

[39] For similar definitions see König (1980: 289); Quirk et al. (1985: 198); Leech (2004: 22); Huddleston and Pullum (2002: 165); Kranich (2013: 18).

The speaker offers his subjective interpretation/explanation of the situation. The progressive's lack of aspectual meaning is reflected in the fact that the A part of the utterance is encoded in the simple aspect.

Smith (2005: 92) makes the important observation that the interpretative progressive frequently occurs in connection with communication verbs. This also applies to the progressive in (3.15).

Pragmatic uses Under this label I subsume all those non-aspectual progressive uses that clearly serve a pragmatic function – for example by expressing emphasis or strong surprise – but that are no *ALWAYS*-type or interpretative progressives. Kranich (2013: 18) refers to them as tentative or emphatic uses, while labels such as highlighting/ prominence or politeness/downtoning are found in Mindt (2000: 249). The progressive in (3.16) expresses a tentative reading, while the one in (3.17) conveys emphasis and surprise.

(3.16) *My dear Henry – I* ***was hoping*** *that by now you were a settled family man and were going to sit down and give us the great fireside books of your later period. You can't start all over again [...].* (Kranich 2013: 18)

(3.17) *What?* ***Am I hearing*** *right?* (Römer 2005: 100)

Recentness progressive Uses in which the past progressive functions as an expression of the recent past and co-occurs with adverbials such as *just*, *recently* or *the other day* were already addressed in chapter 3.3.2. For convenience, the relevant example is presented again.

(3.18) *I* ***was just reading*** *the other day that Billy Graham, as you say he's been an important part of Nixon's life, he officiated at the funeral of Hannah Nixon, who died in 1967* (Pfaff et al. 2013: 219)

Since the process of reading is complete and not viewed from within, the progressive in the example sentence can be said to be perfective in meaning. It seems that apart from expressing recentness such progressive uses are chosen because they convey an informal tone. Thus, their increasing use can be regarded as a sign of colloquialization.

General validity In her analysis of progressive use in spoken BrE, Römer (2005) found a comparatively high number of non-aspectual progressives that do not easily fit any functional category. These uses occur in contexts in which we would normally expect the simple aspect. As the progressives in (3.19) and (3.20), they refer to situations that can be regarded as being generally valid.

(3.19) *But in general the people who* ***are doing*** *some of those things, a lot of it's probably escapism.* (Römer 2005: 96)

(3.20) *If the man and a woman **is working** it's still the woman who does most of the housework.* (Römer 2005: 96)

These use are puzzling from a semantic and pragmatic perspective. Maybe they are just a reflection of an increased readiness to use the progressive also in those contexts in which the simple aspect is the default option and in which no specific functional motivation for the use of the construction can be identified.

3.4.3. Derived aspectual/future uses

Futurate progressive Progressives that refer to planned (near) future events are often called futurate progressives. They can be regarded as derived aspectual uses since their meaning is linked to progressive aspectuality. Kranich explains: "the 'near future' use of the progressive often denotes a situation which is firmly planned or may already be conceptualized as in progress, e.g., because preparatory activities are already ongoing" (2013: 15). Leech et al. (2009) present an example (3.21) in which such a link is evident.

(3.21) *Frank straightened up his desk and went back out through the reception area. I'm **going** to the ranch, he said.* (Leech et al. 2009: 133)

In the light of such examples, Leech et al. claim that "the futurate could be seen as a metonymic extension of the basic meaning of the progressive" (2009: 133).

A futurate progressive with a less straightforward link to the progressive's basic aspectual meaning is the following:

(3.22) *Martin **is coming** over for lunch on Sunday.* (Leech 2004: 33)

Matter-of-course future Progressives of the structure *will/shall* + progressive frequently do not present the future happening as being in progress but in its entirety (i.e. as perfective) and as happening as a matter of course (cf. Celle and Smith 2010; Leech 2004: 66ff.). Kranich classifies these uses as derived aspectual ones (2013: 15ff.) even though such a link is less evident than for futurate progressives. An aspectual reading can hardly be attested for the progressives in (3.23) and (3.24).

(3.23) *I'll **be driving** into London next week.* (Leech 2004: 67)

(3.24) ***Will you be going** to the shops this afternoon?* (Huddleston and Pullum 2002: 171)

3.4.4. The progressive as a lexical-grammatical phenomenon

So far, a central aspect of the progressive's use has not received sufficient attention: the fact that the progressive is by no means a purely grammatical but a lexical-grammatical phenomenon. However, some hints have been provided. For example, I mentioned Smith's observation that interpretative progressives frequently occur in connection with

communication verbs (2005: 92). The recentness progressive's co-occurrence with adverbials such as *just*, *recently* or *the other day* has also been addressed (cf. chapter 3.4.2). *ALWAYS*-type progressives are even labelled according to the lexical items they co-occur with. A further example of lexically specific progressive use is the *BE being adj.* pattern (cf. chapter 3.3.2). Here the adjective frequently is one with a negative connotation, such as *naughty* in example (3.8), referring to the violation of a norm or implying insincerity (Levin 2013: 193).

A very thorough study of such lexical aspects of progressive use is Römer (2005). Studying spoken BrE, she analyses more than nine thousand progressive tokens and comes to the conclusion that "it is difficult, maybe even impossible, to treat the progressive as a grammatical construction independent of lexis" (2005: 169).

Different progressive functions exhibit distinct lexical preferences. For example, tentative, pragmatic uses[40] frequently occur with the verbs *hope*, *think* or *suggest*. By far the most common subject is the personal pronoun *I*; a frequently co-occurring adverbial is *just*. This leads Römer to conclude that "we cannot say that politeness is a function of the progressive in general. Its use in spoken English is to a large extent lexically determined, meaning that only a restricted set of forms can express this function" (2005: 98). The same applies to other functions such as emphatic/attitudinal progressive uses (2005: 99ff.) or uses referring to habits (2005: 103f.).

Focusing on progressive verbs, Römer shows that they exhibit distinct tense form distributions. For example, progressive *happen* has a clear preference for the present tense, while most cases of progressive *wonder* occur in the past tense (2005: 118ff., Table 21). Such verb-specific preferences are also observed for other co-occurrence phenomena such as subjects, objects, adverbials and negation.

Studying the use of the futurate progressive, Nesselhauf and Römer (2007) find that it frequently occurs with the personal pronouns *I* and *we* in subject position, while *it* is extremely rare. The function also shows a tendency to take negated forms, while adverbial co-occurrence is less common than expected. The authors identify relatively general patterns such as *I'm not -ing* and more specific ones such as *I'm not telling you* (2007: 327).

Römer's (2005) findings clearly show that specific progressive functions exhibit a distinct lexical and morphosyntactic profile. Uses such as the *ALWAYS*-type progressive show that this reasoning also works the other way round: lexically specific progressive patterns – in this case present progressives modified by an *ALWAYS*-type adverbial – can acquire distinct pragmatic meanings that become conventionalised discourse strategies over time. Such a development is obviously not restricted to lexically specific progressive patterns but is a general phenomenon. Bybee explains that "constructions can take on special pragmatic and semantic values through repetition despite not being of especially high frequency. The special connotations of the construction [...] have to be acquired by repetition in the appropriate contexts" (2007: 17).[41]

Römer's data strongly suggests that apart from the abstract *be + -ing* participle

[40] Römer refers to these uses with the label "politeness or softening".
[41] See Kay and Fillmore (1999) for a detailed case study.

construction many lower-level patterns exist that guide and restrict speakers' choices when using a progressive. Her observation that "some verbs indeed choose particular progressive structures while others prefer different ones and that, on the other hand, not every verb can be found in every construction" (2005: 170) is very much in line with the reasoning of usage-based Construction Grammar, which assumes that constructions are mentally represented at different levels of abstractness and that more concrete, lower-level constructions are associated with particular lexical items.[42]

It is exactly this issue of mental representation that leads to the question of whether speakers actually memorise some patterns as coherent linguistic units (i.e. as chunks). If yes, such patterns would have acquired the status of constructions (in the sense of Construction Grammar), existing alongside the abstract *be + -ing* participle construction. While such a development is conceivable for sufficiently frequent patterns that have developed a distinct and at least partly non-compositional meaning, it is unlikely to apply to less frequent, fully compositional patterns without a distinct semantic/pragmatic profile.

Finally, it is worth asking what role lexically specific patterns play in the progressive's recent development. Is it possible to identify distinct patterns of use that show a clear diachronic trend? If yes, how do they contribute to the progressive's functional and frequency development?

When researching the progressive's use and development, it is important not only to focus on the "big picture" (by analysing broad and inclusive categories) but also to ask whether meaningful insights can be obtained by studying the more specific, lexical-grammatical uses of the construction.

3.5. Summary and resulting hypotheses

3.5.1. Evaluating the state of research

The present chapter has shown that the scope of research on the progressive is impressive. Nonetheless, there are certain limitations that have to be addressed.

Data base: Advances in corpus development

First, there is the issue of corpus development. Early studies such as Mossé (1938) or Scheffer (1975) provide quantitative findings for pre-digital written texts. Major advances in corpus linguistics happened towards the end of the century, most notably with the compilation of the Brown family corpora, which enabled real-time investigations of the progressive's recent development. Resulting studies, such as Mair and Hundt (1995), revealed that the progressive's frequency development was ongoing in late 20th-century English. In the course of the first decade of the 21st century, further diachronic written corpora were compiled (for example the *Time* Corpus, COHA, and new versions of ARCHER). Based on these corpora, several studies on the progressive's development have been published (e.g. Hundt 2004a, Kranich 2010, Levin 2013).

[42] See Diessel (2016) for an informative overview article.

Due to unavailability of diachronic spoken corpora, the study of the progressive's development in spoken language has long remained a problem. Smith (2005) and Leech et al. (2009) used self-compiled mini-corpora, which produced interesting but – inevitably – inconclusive results. The first diachronic spoken corpus of BrE, DCPSE, was released in 2006. Aarts et al. (2010) used it to quantify the progressive's frequency development in late 20th-century spoken BrE.

Unlike the recently published and extremely large *Hansard* corpus (Alexander and Davies 2015), which contains official transcripts of UK parliamentary debates, DCPSE contains very tidy, high-quality data, enabling genre-specific investigations.[43] Importantly, the largest part of the DCPSE data comes from the most natural of all genres – spontaneous conversation. Unfortunately, though, the most recent data is now already 25 years old. Thus, there is a one-generation gap between DCPSE and present-day language use. In the light of studies that show that the progressive's recent development proceeded very swiftly (e.g. Smith 2005, Leech et al. 2009), it has to be assumed that DCPSE no longer reflects the progressive's current state of use. This is why I decided to compile a new corpus of truly present-day spoken BrE, designed to match the DCPSE as closely as possible. This new corpus resource – the *Freiburg Corpus of Spoken English* (FCSE) – supplements the two existing subcorpora of DCPSE – LLC (data from 1950-70s) and ICE-GB (data from 1990-2) – by providing a third subcorpus with data from 2012-16.[44] The corpus will enable statements regarding the progressive's ongoing development in different genres of spoken BrE. Furthermore, the data will clarify whether developments in the progressive's spoken use observed for the late 20th century were mere fluctuations or part of a real trend.

Apart from five other genres, FCSE contains 150,000 words of transcribed informal spontaneous speech. While this certainly constitutes a very valuable resource, it can by no means rival the new spoken *British National Corpus* (Spoken BNC2014), which was being compiled and finally released at the time of writing (Love et al. 2017). The Spoken BNC2014 comprises more than eleven million words of exclusively spontaneous conversation and, in combination with the existing BNC (containing data from the 1990s), provides an invaluable resource for the study of language change in progress.[45] However, as previous research has shown, a central aspect of the progressive's development is its genre-specific nature. In this context – and regardless of the high value of Spoken BNC2014 – it was important to compile a new corpus of spoken BrE that enables the researcher to take into account exactly this aspect of the progressive's development and use. Furthermore, the fact that Spoken BNC2014 was only released in late 2017 (i.e. in the late stages of this research project) made it imperative to rely on self-transcribed spontaneous conversation.

[43] See chapter 4.2.1 for a detailed description of DCPSE.
[44] See chapter 4.2.2 for a detailed description of the new corpus.
[45] See chapter 4.2.3 for a detailed description of the two corpora.

Theoretical foundation and methodology

Apart from the focus on 20th-century written English, a second limitation of the existing body of research has to be addressed. It concerns the theoretical orientation of previous studies as well as their methodology.

Regarding the theoretical orientation, all studies – by focusing on corpus data and making reference to concepts such as grammaticalization or colloquialization – adopt a usage-based perspective. In this sense, many of the cited studies are similar to Mair and Hundt (1995). However, as has been outlined in chapter 2, the field of usage-based linguistics has rapidly evolved in the course of the past 25 years, providing new concepts and insights – for example on frequency effects. Equally important, CxG has now become the pervasive theory in usage-based approaches.

While most of the studies on the progressive's development and use are generally compatible with these developments, they do not put them centre stage but stick to the theoretical frame established by Mair and Hundt in 1995. One notable exception is Römer (2005) (cf. chapter 3.4.4), who studies the progressive as a lexical-grammatical phenomenon, thereby acknowledging the insight from UBL that lexis and grammar are closely connected.

Furthermore, the statistical methods used in the majority of studies do not really go beyond what Mair and Hundt did as early as 1995, meaning that the quantitative analyses are often descriptive in nature. If inferential methods are applied, they normally remain on the level of basic significance testing.[46] While this criticism shall by no means be understood as dismissive or conceited, it is inevitable in the light of recent developments in corpus linguistic methodology. For example, typicality measures such as Collostructional Analysis (Stefanowitsch and Gries 2003; Gries and Stefanowitsch 2004) or Hierarchical Configural Frequency Analysis (Gries 2009: 248ff.; Hilpert 2013a: 55ff.) can provide very fine-grained insights, going beyond simple relative frequencies. Regression modelling (Gries 2013b: 247ff.) can be a very helpful tool in the analysis of complex lexical-grammatical phenomena.

It is exactly such a usage-based CxG perspective combined with different analytical tools offered by quantitative corpus linguistics that is adopted in this study. Thereby, I aim to provide an innovative approach towards the progressive's contemporary use and recent development that is built on, but goes beyond the existing body of research.

3.5.2. Research questions and hypotheses

I will conclude this chapter by formulating my central hypotheses regarding the progressive's ongoing development and use in spoken BrE.

[46] Exceptions can be found among studies of the progressive's use in world Englishes. A recent example is Deshors (2017), who uses a combined collostructional- and correspondence-analysis approach. Also studies from the field of learner corpus research have made use of complex statistical tools in the analysis of the progressive; e.g. Rautionaho and Deshors (2018).

Frequency development in 21st-century spoken English

In the light of the results for the progressive's frequency development in spoken 20th-century English (cf. chapter 3.3), I assume that the construction's frequency increase is likely to continue in the 21st century:

Hypothesis 1 The frequency increase of the progressive is still under way in the 21st century, and it is proceeding from different baselines at different speeds in different genres.

As for the driving forces of the frequency increase, I base my argumentation on Leech et al. (2009), who found that it is in the already established contexts of use in which the progressive is developing most rapidly (cf. chapter 3.3.2). Even though I also expect the progressive to become more frequent in contexts that have so far not favoured its use, I do not expect use in these innovative contexts to be the major driving force of the frequency increase. Based on these assumptions, the following hypothesis is formulated:

Hypothesis 2 The frequency increase of the progressive in the 20th and 21st centuries is based on further expansion of prototypical core uses as well as on structural or functional innovations. Structural and functional innovations are expected to have less impact on the corpus data than the development of the prototypical core uses.

Studying lexical-grammatical progressive patterns

I have argued that it is important to analyse the progressive on the level of lexically specific patterns of use. In addition to identifying and quantifying the use of such patterns, it is also vital to analyse the contexts they occur in and to closely study their functions in discourse. In line with research on frequency effects, I assume that repeatedly occurring progressive patterns develop specific pragmatic meanings, become part of speakers' routinised discourse strategies and "are registered in linguistic memory indexed with their implications and contexts of use" (Bybee 2006: 721).[47] Hence, I formulate the following hypothesis:

Hypothesis 3 The progressive has to be analysed at different levels of complexity and abstraction (structurally as well as functionally). Frequent use of lexically specific progressive patterns in specific contexts is linked to the development of distinct pragmatic meanings and to increasing routinisation.

[47] See also Bybee (2015: 133ff.).

4. Data & methodology

The present chapter presents the data and methodology used in this study. Its first part is dedicated to the corpus-based approach, followed by the introduction of a new corpus of spoken British English, which has been compiled by the author. Furthermore, the data retrieval procedure and different techniques for data analysis are addressed.

4.1. The corpus-based approach

As the previous chapter has shown, a broad range of corpus resources has been used to study the progressive's diachronic development. Offering authentic language data in digital format, corpora are a primary source for the study of grammatical variation and change. Thus, it seems appropriate to briefly address a few important aspects of the corpus-based approach before introducing the corpora used in the present study. However, since the body of literature concerned with theoretical and methodological questions of corpus linguistics is truly immense,[1] the present section can only address selected aspects that are considered particularly relevant.

It is obvious that by studying authentic usage data, corpus linguistics is closely related to the usage-based approach (cf. Leech 2000: 685ff.). However, among corpus linguists, it is not unanimously agreed upon how exactly corpus data should be used to inform usage-based descriptions and theories of language. Broadly speaking, corpus linguistic studies can be situated on a continuum between *corpus-based* approaches at one end, and *corpus-driven* approaches at the other end (cf. Xiao 2009: 993ff.). Xiao explains it the following way:

> corpus-driven linguists aim to build theory "from scratch" claiming that they are completely free from pre-corpus theoretical premises and base their theories exclusively on corpus data, [...] whilst corpus-based linguists tend to approach corpus data "from the perspective of moderate corpus-external premises" [...] with the aim of testing and improving such theories. (2009: 993)

In practice, however, basically every corpus study is to varying degrees informed by theoretical assumptions, meaning that true corpus-driven studies are very rare. As the title of this subchapter already suggests, the present study considers itself to be closer to the corpus-based than to the corpus-driven end of the continuum, since it explicitly

[1] An impressive range of rather general topics is addressed in the following handbooks, which have all been published within the past ten years: Biber and Reppen (2015); O'Keeffe and McCarthy (2010); Lüdeling and Kytö (2008 & 2009).

acknowledges previous research on the progressive as well as the assumptions of usage-based CxG.

Following these general remarks, I briefly want to address the role of corpora in the study of recent and ongoing language change and in the study of spoken language.

The study of recent and ongoing – especially lexical and grammatical – change has been a major focus of English corpus linguistics in the past twenty years and has resulted in publications such as Hundt and Mair (1999), Mair (2006b), Leech et al. (2009), Leech et al. (2012), Aarts et al. (2013), to name just a few. This interest goes hand in hand with the development of suitable corpus resources, as Leech explains:

> a diachronic interest in PDE has developed through the compilation of corpora enabling a precise study of changes in English over the recent past: for example, the Brown family of corpora, Bas Aarts's *Diachronic Corpus of Present-day Spoken English* (DCPSE) and Mark Davie's *Corpus of Contemporary American English* (COCA) and, with a longer time span, his *Corpus of Historical American English* (COHA). [...] [C]ontemporary language [...] is no longer seen as a synchronic entity, but as an entity continually subject to change, even within the time-span of a decade or two. (2015: 159)

While it was sociolinguistics that pioneered the the study of ongoing change, corpus linguistics now appears to be the mainstay of the field. Real-time research based on corpora is certainly easier than complex longitudinal sociolinguistic community studies, which take up a lot of time and resources (cf. Mair 2009: 1110ff.). This, however, does not mean that sociolinguistic insights have no role to play in the study of ongoing change. On the contrary, concepts such as *colloquialization* (Mair 2006b) are of central importance to the meaningful interpretation of corpus frequencies. This point is encapsulated in the following quote by Mair, which can be read as a plea for a corpus-based as opposed to a corpus-driven approach to the study of ongoing change: "Corpus-based empiricism, however, will lead to nothing more than the accumulation of under-analysed and frequently pointless statistics unless the interpretation of the results is carried out in an appropriate theoretical framework" (2009: 1122).

The continuous development of new corpus resources has also resulted in the compilation of spoken corpora or corpora containing at least a substantial part of spoken data (see also chapter 3.5.1). For British English, examples are the *London-Lund Corpus* (LLC), the *British National Corpus* (BNC), or the *International Corpus of English – Great Britain* (ICE-GB). Since spoken data is generally considered "the site of origin of almost all non-prestige innovations in language" (Mair 2009: 1110), these corpora are of particular interest to the linguist interested in present-day English and its recent developments. One of the corpora used in the present work – the *Diachronic Corpus of Present-day Spoken English* (DCPSE) – was specifically designed for the study of recent grammatical change. In addition to such small corpora as the DCPSE, much larger corpora such as the 1.6 billion words *Hansard* corpus (containing the official transcripts of UK Parliamentary debates) are now available. As already mentioned in chapter 3.5.1, another major spoken corpus, the Spoken BNC2014, was being compiled and released at Lancaster University at the time of writing (Love et al. 2017).

With regard to this increasing availability of different corpus resources, Mair (2006a) argues for methodological pluralism, encouraging the use of both carefully compiled small corpora and less carefully collected large corpora. The former come with a lot of metadata and well-balanced material. However, they are often too small for the analysis of specific, low frequency phenomena. Compared to that, larger corpora offer huge amounts of data, which is often not perfectly well documented and balanced. According to Mair, corpus linguists should embrace the possibilities offered by both kind of resources and use them in complementary fashion. He argues that "restricting the scope of one's work to data available in a small number of corpora only would be counter-productive in the analysis of many linguistic phenomena" (2006a: 370).

In a comprehensive review article, Leech (2000) discusses how spoken corpus research has influenced the conception of English grammar. Since the spoken medium is the primary medium of language use, he regards the analysis of spoken corpora as crucial for the construction of usage-based theories of language (2000: 685ff.). While he rejects a strict dichotomy of written and spoken English grammar (2000: 687ff.), he emphasises the findings of spoken corpus research. For example, private face-to-face conversation is characterised by a shared context of the interlocutors, leading to high frequencies of linguistic features that reduce the length and complexity of utterances (e.g. personal pronouns, types of ellipsis, non-clausal material, etc.) (2000: 694ff.). Furthermore, conversational grammars are very interactive, meaning that questions, imperatives and first- and second-person pronouns figure prominently. They also reflect speakers' tendencies to interact through contrastive perspectives (frequent use of negatives and the conjunction *but*) (2000: 696). Other features of conversation are attitude/stance markers (e.g. stance adverbials or interjections) and a highly repetitive lexico-grammatical repertoire (recurrent word sequences are much more common in conversation than in writing) (2000: 697). An important work in this regard is the entirely corpus-based *Longman Grammar of Spoken and Written English* (Biber et al. 1999), which identifies many features characteristic of spoken conversation. It also shows that the progressive is much more common in conversation than in written genres (1999: 462).

Apart from the continuous development of corpus resources, also corpus-linguistic methodology is continuously being improved (cf. chapter 2.3 & 3.5.1). Many sophisticated statistical tools are publicly available, often even for free. One example is the programming language *R* (R Core Team 2017), which offers many corpus tools on the one hand, and many statistical packages on the other hand, making it an extremely powerful tool for the compilation, exploitation and analysis of corpora.[2] Techniques such as collostructional analysis or regression modelling (see chapter 4.4) have all been implemented in *R*. Today, statistical modelling of grammatical phenomena, i.e. their corpus frequencies, has become an integral part of corpus linguistics. This is acknowledged by Leech, who states that "corpus-based statistical models have advanced the study of grammar" (2015: 155).

A further development in corpus linguistics derives from the fact that usage-based theories of language aim at psychological plausibility (cf. chapter 2). Thus, it is not

[2]Instructive introductions are Gries (2013b & 2017a), Baayen (2008), and Field et al. (2012).

surprising that studies combining corpus and elicitation data have become more frequent in recent years (e.g. Gries et al. 2005; Gries et al. 2010; Ellis et al. 2016; Horch 2017). Similarly, the question if and in how far corpus data can reveal explicitly cognitive/psycholinguistic insights has recently received considerable attention (e.g. Schmid 2017b; Gries 2017b; Dąbrowska 2016). What is more, a new subfield seems to be forming: cognitive historical linguistics (cf. Mair 2017 in Hundt et al. 2017, and also Petré 2017). It is based on the assumption that the study of language history and change can – at least to some degree – be informed by cognitive and psycholinguistic insights in mechanisms such as chunking (Ellis 2017; Bybee and Moder 2017), analogy (Behrens 2017; De Smet and Fischer 2017), or priming (Pickering and Garrod 2017; Mair 2017) (cf. also chapter 2.1.1).

It is the aim of the present work to do justice to these developments in corpus resources, statistics and theory and to provide a state-of-the-art analysis of the progressive's use and development in present-day spoken English.

I will now introduce the corpora used in the present study.

4.2. Corpus resources

4.2.1. The *Diachronic Corpus of Present-Day Spoken English* (DCPSE)

The DCPSE was compiled by the Survey of English Usage at University College London and released in 2006 (Aarts and Wallis 2006).[3] Its data comes from two existing corpora of spoken British English: the *London-Lund Corpus* (LLC) and the spoken part of the *International Corpus of English – Great Britain* (ICE-GB). The LLC part comprises texts from the 1950s (3 texts à ca. 5000 words), 1960s (36 texts à ca. 5000 words), and 1970s (52 texts à ca. 5000 words), while the ICE-GB texts range from 1990 to 1992 (212 texts à ca. 2000 words). Altogether, each subcorpus comprises more than 400,000 words of spoken data, amounting to an overall corpus size of 885,436 words. DCPSE is tagged and parsed, enabling the analysis of complex grammatical and syntactic phenomena. It comes with the corpus software ICECUP (Nelson et al. 2002), which, among many other search options, offers so-called *Fuzzy Tree Fragments* (FTFs) to search directly for syntactic labels and structures. DCPSE is annotated for a range of variables, such as genre, speaker age and gender, etc. While the text sizes of the two subcorpora differ, most of the genres are of comparable size (see Table 4.1).

DCPSE was specifically designed for the research of recent grammatical variation and change and, thus, is very well suited to study the progressive's genre-specific use and development in late 20th-century spoken British English. Since the progressive is highly frequent, the comparatively small size of the DCPSE does not pose a problem with regard to the construction's major developments. Only when it comes to very specific progressive uses/patterns, the use of supplementary corpus material (for example from BNC) becomes necessary.

However, as already mentioned in chapter 3.5.1, DCPSE is no longer up to date for

[3] For reviews see: Hilpert (2010); O'Donnell (2008); Davies (2009).

Genre	Words		
	LLC	ICE-GB	Sum
Informal Face-to-face Conversations	218,307	185,537	403,844
Formal Face-to-face Conversations	48,896	41,879	90,775
Telephone Conversations	26,828	20,414	47,242
Broadcast Discussions	45,237	43,920	89,157
Broadcast Interviews	20,899	22,147	43,046
Spontaneous Commentary	46,426	48,955	95,381
Parliamentary Language	10,494	10,589	21,083
Legal Cross-examination	5,277	4,381	9,658
Assorted Spontaneous	10,528	11,147	21,675
Prepared Speech	31,182	32,393	63,575
Sum	464,074	421,362	885,436

Table 4.1.: Genre sizes DCPSE.

the study of ongoing language change. At the time of writing, the ICE-GB data was already more than 25 years old, meaning that it cannot shed light on the progressive's frequency development in the 21st century. In order to analyse the progressive's ongoing development in different genres, I compiled a new corpus of spoken British English that roughly matches the DCPSE and that extends its diachronic scope.

4.2.2. A new corpus of spoken English: Extending the diachronic scope of the DCPSE

The aim was to compile a corpus of spoken British English containing data from 2012 to 2016, mirroring the major genres of the DCPSE and extending its diachronic scope by ca. 25 years. However, the compilation of spoken corpora is time-consuming and comes with a number of challenges.[4] Suitable transcripts of spoken language are rare and often of low quality (e.g. inconsistent transcription of specific features, transcription errors, adjustments to written norms, etc.). In order to ensure feasibility of the project, it was decided to reduce the number of genres to six: Three conversational genres (*Informal Face-to-face Conversation*, *Broadcast Discussions*, *Broadcast Interviews*) and three non/less-conversational genres (*Prepared Speech*, *Parliamentary Language*, *Spontaneous Commentary*). For each of these genres, it was either possible to find existing transcripts on the internet (which had to be corrected, however), or suitable recordings that had to be transcribed. This hybrid approach of using and correcting existing transcripts on the one hand, and of transcribing suitable recordings on the other hand, considerably reduced the workload compared to an approach in which no existing transcripts had

[4] A considerable number of publications address the compilation of spoken corpora. A very useful one is Thompson (2005). Others include: Adolphs and Knight (2010); Adolphs and Carter (2013); Ruhi (2014); McCarthy and O'Keeffe (2009); Kirk and Andersen (2016); Wichmann (2008).

been used. Given the fact the corpus compilation was part of a single PhD project, the hybrid approach turned out to be the only viable one. Table 4.2 illustrates the design of the corpus, which will subsequently be called *Freiburg Corpus of Spoken English* (FCSE).

Genre	Data source	words
Informal Face-to-face Conversations	existing recordings	149,690
Broadcast Discussions	existing transcripts	55,930
Broadcast Interviews	existing transcripts	52,005
Parliamentary Language	existing transcripts	52,051
Prepared Speech	existing transcripts	102,762
Spontaneous Commentary	existing recordings	47,540
Sum		459,978

Table 4.2.: Genre sizes *Freiburg Corpus of Spoken English* (FCSE).

It has to be noted that FCSE's genre sizes differ from those of DCPSE. First, this is due to the fact that certain genres of DCPSE – e.g. *Broadcast Interviews* (ca. 20,000 words) or *Parliamentary Language* (ca. 10,000 words) – are extremely small. In these cases, it was decided that more robust numbers are more important than easy comparability without normalisation. Second, certain genres have already got different word numbers in DCPSE. This is for example the case for *Informal Face-to-face Conversation*, which comprises ca. 33,000 more words in LLC compared to ICE-GB. Thus, progressive numbers needed to be normalised anyway. Third, genre sizes were also influenced by the availability of suitable material. For *Prepared Speech*, for example, it was comparatively easy to find a large number of transcripts that needed relatively little correction. Thus, it was possible to achieve a genre size of ca. 100,000 words. For *Broadcast Interviews*, a large number of transcripts could be obtained as well. However, the corresponding recordings were not always available. Thus, with ca. 50,000 words, the genre is considerably smaller than *Prepared Speech*. It is, however, still much larger than *Broadcast Interviews* in LLC and ICE-GB, which – as already mentioned – only comprise ca. 20,000 words each. For *Spontaneous Commentary* and *Informal Face-to-face Conversation*, it was the transcription process that proved to be a limiting factor. While it was possible to achieve the same word number as in LLC and ICE-GB for *Spontaneous Commentary*, time and monetary resources only permitted a size of ca. 150,000 words for *Informal Face-to-face Conversation*.[5]

The first step in the compilation process was the data collection. The internet was searched for suitable transcripts dating from 2012 to 2016. If freely available transcripts were found, it was first checked whether the corresponding recordings were available. If this was not the case, the transcripts were discarded. If transcripts were available together with the recordings, their quality was evaluated against the recordings. In many cases (for example transcripts of university lectures), the transcripts were rather

[5]In the present work, these differences in genre size do not pose a problem since I always present the normalised progressive frequencies alongside the absolute numbers (cf. chapter 4.4.1).

rough content-focused paraphrases than faithful linguistic representations. All these transcripts were discarded as well. For four genres, however, it was possible to obtain a large number of transcripts that constitute comparatively faithful orthographic transcriptions. As indicated in Table 4.2, these are *Parliamentary Language*, *Prepared Speech*, *Broadcast Interviews*, and *Broadcast Discussions*. No suitable transcripts could be found for *Informal Face-to-face Conversation* and *Spontaneous Commentary*.

For parliamentary debates, the collection of transcripts was straightforward: The UK House of Commons Hansard archive is publicly accessible and contains thousands of recorded debates and their official transcripts. Since this material has already been made available in the *Hansard* corpus (Alexander and Davies 2015), collecting a much smaller sample of these transcripts and including them in a new corpus might look like a futile effort. However, the Hansard transcripts are by no means entirely faithful renditions of the House of Commons debates. As Mollin already pointed out in 2007:

> the transcripts omit performance characteristics of spoken language, such as incomplete utterances or hesitations, as well as any type of extrafactual, contextual talk (e.g. about turn-taking). Moreover, however, the transcribers and editors also alter speakers' lexical and grammatical choices towards more conservative and formal variants. Linguists ought, therefore, to be cautious in their use of the Hansard transcripts and, generally, in the use of transcriptions that have not been made for linguistic purposes. (2007: 187)

A comparison of the transcripts to their recordings confirmed Mollin's observation. Consequently, it was decided to sample ten transcripts à 5000 words[6] and to check and correct them against the actual recordings. This way, the transcripts' quality could be seriously improved in a relatively time efficient way.

A similar approach was adopted for the remaining three genres that are based on existing transcripts. For *Prepared Speech*, transcripts of eight political/public speeches, four sermons, and twelve public lectures were collected. They were retrieved from web pages of universities, churches, political parties, etc. All of these transcripts were thoroughly checked and corrected against the corresponding recordings.

Transcripts of *Broadcast Interviews* come from the BBC Radio 4 programme *In Touch* and from the BBC TV programme *The Andrew Marr Show*, a weekly political interview show. Both shows' internet archives contain hundreds of recordings and their corresponding transcripts. A total of 25 transcripts were sampled, checked and corrected. More than half of the *Andrew Marr Show* transcripts do not feature Andrew Marr but different stand-in moderators as the shows' host. These transcripts were deliberately chosen in order to avoid an extreme overrepresentation of one speaker (Andrew Marr) in the data.

For *Broadcast Discussions*, transcripts of the more involved BBC radio talkshow *Ouch* proved to be well-suited. While *In Touch* and *The Andrew Marr Show* represent prototypical interviews, the *Ouch* talkshow format entails a higher number of speakers and more open discussion than closed questions. Fortunately, the show has had changing

[6] A list of all corpus texts is provided in the appendix.

moderators, meaning that none of them is unduly overrepresented in the data. Again, all transcripts were compared to the recordings and manually corrected.

No suitable transcripts could be found for the *Spontaneous Commentary* genre. Thus, recordings of sports (seven texts) and ceremonial commentary (three texts) were transcribed orthographically. The transcripts include speech-specific features such as pauses, overlaps, hesitations, contractions, etc. Information about interruptions, laughs or coughs was also included, and spelling variants were used consistently. While this procedure resulted in transcripts of comparatively high quality, it was, however, much more time-consuming than the post-editing of already existing transcripts.

Compilation of the last genre, *Informal Face-to-face Conversation*, proved most time-intensive. Above all, good recordings of authentic face-to-face conversation are rare. Fortunately, though, the British Library and the BBC have set up a project collecting and archiving exactly such conversations. It is called the *Listening Project* and has been running since 2012, collecting hundreds of conversations between friends and close relatives. On the project website it is described as follows:

> The Listening Project is an audio archive of conversations recorded by the BBC. People are invited to share an intimate conversation with a close friend or relative, to be recorded and broadcast [...] by the BBC and curated and archived in full by the British Library. These one-to-one conversations, lasting up to an hour and taking a topic of the speakers' choice, collectively form a picture of our lives and relationships today. [...] [T]he participants are free to talk about what they want and to direct the conversation themselves. (http://sounds.bl.uk/Oral-history/The-Listening-Project; accessed: 2017-08-08)

Even though the project was set up as an oral history project, the conversations of the *Listening Project* are a very valuable linguistic resource. After listening to a large number of them, a total of 19 recordings were selected for transcription. Age and gender balance as well as regional variation of the speakers was taken into account as far as possible. The transcription process followed exactly the same rules as the transcription of the *Spontaneous Commentary* recordings. Almost 150,000 words were transcribed accordingly.

In the course of the data collection, it was also checked whether all speakers are native speakers of British English. For certain genres, this could easily be found out. Members of parliament, for example, can be expected to have English as their native language or at least as one of several native languages. In case of doubt, one can always check their Wikipedia entry and/or a personal internet page and read through the biographical information. Also the speakers of the *Listening Project* did not pose a problem. First, the project's metadata provides a certain amount of background information. Second, the speakers normally give away a lot of personal information in the course of their conversations, making it possible to infer central parts of their biography. For the other genres, however, unambiguous identification could be difficult. In controversial cases, I relied on three cues: First, close listening to characteristic and easy-to-identify accent features such as rhoticity – to exclude, for example, American English speakers. Second,

searching the transcript for biographical information, and third, searching the internet for information about the speakers. If these efforts did not result in an unambiguous judgement, the respective transcripts were not included in the corpus.

Finally, all transcripts underwent a standardisation procedure, in the course of which all metadata expressions, abbreviations, acronyms, quotation marks, etc. were adjusted according to a self-written standard. Following this standardisation, the transcripts were processed in a *WebLicht*[7] tool chain, including tokenisation, lemmatisation and POS tagging. The *WebLicht* output underwent final processing and merging in *R*.

Altogether, FCSE comprises roughly 460,000 words of spoken British English recorded between 2012 and 2016. While the orthographic transcriptions are not well-suited for the study of prosodic or phonetic features, and corpus size is not large enough for the study of many lexical phenomena, the corpus is certainly a valuable resource for the study of frequent grammatical constructions. Together with the DCPSE, it covers a time span of ca. 55 years and comprises ca. 1.35 million words – thus representing a unique resource for the study of recent and ongoing grammatical change in different genres of spoken British English.[8]

4.2.3. BNC and Spoken BNC2014

In addition to DCSPE and FCSE, the present study also makes use of the conversational part of BNC (i.e. the demographically sampled section of the spoken component; henceforth called BNC-DS) (Crowdy 1995) and of the recently released Spoken BNC2014 (Love et al. 2017). While DCPSE and FCSE represent the two major corpus resources for the study of the progressive's recent genre-specific frequency development, BNC-DS and Spoken BNC2014 are mainly used to study lexical-grammatical progressive patterns.

The Spoken BNC2014 contains ca. 11.5 million words of informal spoken conversation and was designed as an update of the demographically sampled part of the original BNC. While the BNC-DS data was collected in the early 1990s, Spoken BNC2014 comprises data collected between 2012 and 2016. Its major advantage compared to FCSE clearly is its enormous size, which allows the analysis of lexically specific progressive patterns that are comparatively infrequent. Since the present study seeks to illuminate the progressive's recent diachronic development, these patterns will not only be studied in Spoken BNC2014 but also in BNC-DS.[9]

In the face of the release of Spoken BNC2014 one might wonder why the present work uses it only for the study of lexical-grammatical progressive patterns but not for a more

[7] *WebLicht* is "an execution environment for automatic annotation of text corpora. Linguistic tools such as tokenizers, part of speech taggers, and parsers are encapsulated as web services, which can be combined by the user into custom processing chains. The resulting annotations can then be visualized in an appropriate way, such as in a table or tree format" (weblicht.sfs.uni-tuebingen.de; accessed: 2017-08-08). *WebLicht* was developed as part of the CLARIN-D project and represents one of the most powerful, flexible, and easy to use corpus building tools currently available.

[8] Unfortunately, though, due to copyright issues, the corpus is currently not publicly available.

[9] The present work used the BNC-XML version provided by Lancaster University's CQPweb server (https://cqpweb.lancs.ac.uk) (Hoffmann and Evert 2006). In this version, the demographically sampled part of BNC comprises ca. 5.0 million words.

comprehensive account of the construction's recent frequency development in conversation. First, as already argued in chapter 3.5.1, the progressive's frequency development needs to be analysed in different genres. While the design of DCPSE and FCSE licenses such a genre-specific approach, a diachronic comparison of BNC-DS and Spoken BNC2014 does not. Still, one might object that the 150,000 words of conversation in FCSE could have been replaced with the much larger Spoken BNC2014 material. This, however, would have rendered FCSE an incomplete – and, arguably, not very attractive – extension of DCPSE. Furthermore, FCSE was completed – and progressives retrieved – before Spoken BNC2014 was released. It is for these reasons that the analysis of the progressive's frequency development in chapter 5 mainly relies on DCPSE and FCSE.

For the subsequent study of lexically specific progressive patterns, however, use of the Spoken BNC2014 was possible as well as advisable: While pattern identification was undertaken in DCPSE and FCSE, the qualitative and quantitative analysis is based on the much larger output of the two BNC corpora.

4.3. Data retrieval

4.3.1. What counts as a progressive?

Before retrieving all progressive tokens from DCPSE and FCSE, it had to be determined what actually counts as a progressive. Not all instances of *be* followed by a form ending in *-ing* are progressives. This is obvious for examples as the following, in which the finite form of *be* is followed by a noun ending in *-ing*:

(4.1) *This is a lovely ring.*

There are, however, cases that are less obvious. First, there is the so-called *going to*-future (example 4.2), which looks like a progressive and can be historically derived from it. However, the construction is not normally considered a progressive and was excluded from the data.

(4.2) *[...] we are going to see them in Oxford later this year.* (FCSE_FFC_4_271)

In addition to that, participles functioning as adjectives were excluded:

(4.3) *I'm not amazing at listening.* (FCSE_FFC_2_1209)

(4.4) *And it's just so frustrating.* (FCSE_FFC_7_276)

Similarly, gerunds, i.e. verb forms fulfilling a noun-like function, were neither counted as progressives:

(4.5) *So it's about opening our doors to the public.* (FCSE_FFC_4_530)

(4.6) *See that, that's not bullying.* (FCSE_FFC_8_148)

Adverbially used participles also had to be excluded from the counts since they are no progressives proper.

(4.7) *[…] they're out there searching for somebody […].* (FCSE_FFC_2_1462)

(4.8) *And there we are, all dancing together.* (FCSE_FFC_4_558)

Finally, inverted progressives (i.e. a present participle followed by a form of *be*) – which are extremely rare anyway – were also excluded from the analysis:

(4.9) *Making their way south in the opposite direction are groups of Kuwaitis and Egyptians who had been transported to Basra by the Iraqis (,).* (DCPSE:DI-J05)

Included, however, were cases in which *be* refers to more than one participle. In the following example, both *talking* and *telling* were counted as progressives.

(4.10) *[…] you were talking about the future and telling the critics to stop […].* (FCSE_BI_12_42)

4.3.2. Retrieval of progressives from DCPSE and FCSE

Since DCPSE is syntactically parsed, retrieval of progressive tokens does theoretically not pose a problem. However, as the help files of the corpus make clear, different parsers were used for the two subcorpora: "The parsing of the LLC part of DCPSE was carried out in a different manner from that of ICE-GB. [...] a new bottom-up partial parser which was trained on ICE-GB [...] [was] applied to the tagged LLC subcorpus" (Aarts and Wallis 2006).

In the face of this information, it was decided to compare the progressive output of the two subcorpora. Using *Fuzzy Tree Fragments* (FTFs)[10] (the ICECUP search tool for syntactic structures), I obtained the output for both subcorpora and compared it to the output from a small raw text sample of each subcorpus, which was manually searched for progressives.

In the raw text sample of LLC, the manual search found 125 progressive tokens, while the FTF search found only 100 tokens (80% accuracy). For the ICE-GB sample, the FTF search found 114 progressive tokens, compared to 116 manually identified tokens (98.3% accuracy). This means that the parsing of the LLC subcorpus does not reliably identify progressives, while the parsing of the ICE-GB subcorpus only misses very few cases.

On the basis of these results, it was decided to retrieve the progressives from LLC based on lexical wildcard searches. While this entailed a lot of manual post-editing, it ensured that no progressive tokens were missed in the analysis. The LLC subcorpus was searched for all possible forms of *be*, followed by optional intervening material (not

[10] I searched for all auxiliaries marked as being part of a progressive. This yielded slightly more accurate results than a search for all verb phrases marked as progressive.

limited in word number) and a word ending in *-ing*. In addition to the actual hits, this search yielded a large number of tokens that had to be manually excluded.

For ICE-GB, all progressives were retrieved based on the corpus' parsing using FTFs. Even though the procedure is not 100% accurate, a rate of 98.3% seems good enough to justify automatic retrieval and to avoid a lot of time-consuming post-editing.

Finally, progressive retrieval from FCSE was based on regular expression searches in *R*. Similar to the retrieval process in LLC, the corpus was searched for all possible forms of *be*, followed by optional intervening material (no word limit) and a word ending in *-ing*. Again, the output was thoroughly post-edited, eliminating false positives.

Altogether, the DCPSE and FCSE data yielded a total of 8,085 progressive tokens. These were annotated for a number of linguistic variables such as morphosyntactic context or semantic class of the main verb.

4.3.3. Retrieval of progressive patterns

Identification of recurrent lexical-grammatical patterns was based on the DCPSE and FCSE progressive data (the exact procedure is described in chapter 6.3.1). After close examination of the output, certain patterns were selected for further analysis. These, as well as their corresponding simple aspect realisations, were then searched in BNC-DS and Spoken BNC2014. This procedure comes with two major advantages: First, a database of more than 8,000 progressive tokens (i.e. the DCPSE & FCSE output) guaranteed that frequently used progressive patterns could faithfully be identified in a bottom-up fashion. Second, the large amount of spoken data provided by the two BNC corpora allowed a much more thorough analysis of these patterns than DCPSE and FCSE would have permitted.

The retrieval of the patterns from BNC-DS and Spoken BNC2014 was fairly unproblematic. String searches involving regular expressions and part-of-speech tags returned all relevant hits and their simple aspect encodings. In addition to that, the relevant metadata was collected for each token (text-/division-ID and speaker-ID).

4.4. Data analysis

The following sections present the techniques and tools used for data analysis. First, two different frequency metrics and their interpretation are discussed. Afterwards, I introduce two quantitative corpus-linguistic methods: Collostructional Analysis and Hierarchical Configural Frequency Analysis. Finally, logistic regression modelling will be addressed.

4.4.1. Frequency metrics

With regard to frequency metrics, two questions need to be answered: First, how exactly should the progressive's frequency development be quantified? Second, how should individual variables (i.e. aspects of progressive use) be measured in order to explain the observed frequency development?

The great majority of studies has measured the progressive's corpus frequency per million words (pmw) or per hundred thousand words (htw)[11]. These measures come with two obvious advantages: They are easy to calculate and ensure comparability to the results of other studies. There are, however, some linguists who have proposed more refined frequency measures than simple normalisation. First, there is Smitterberg (2005), who advocates his so-called S-coefficient as a suitable means to study progressive frequencies over time. According to him, a truly variationist approach should study choice relative to opportunity. Thus, it is not sufficient to just normalise progressive frequencies against a baseline of words. Instead, the S-coefficient measures the number of progressives against the number of finite verb phrases minus the number of imperative verb phrases and the so-called *going to*-future (2005: 48). Smitterberg argues that this measure is superior to normalisation per million or hundred thousand words since it takes into account that the number of verb phrases – and thus the number of contexts in which the progressive can potentially be chosen – is not diachronically stable. As the number of verb phrases changes, the contexts of opportunity for progressive use change as well. Non-finite verb phrases, imperatives, and *going to*-futures are removed since they exclude the option of using a progressive, as Smitterberg claims. Consequently, he regards the S-coefficient as a more valid tool to measure speakers' choices over time than simple normalised frequency counts.

In their study of progressive use in DCPSE, Aarts et al. (2010) follow Smitterberg's line of argument and – in addition to simple normalised frequencies – use a measure closely related to the S-coefficient (progressives relative to the number of all verb phrases minus imperatives and *going to*-futures). Like Smitterberg, they argue that

> [c]alculating the frequency of the progressive [...] [per hundred thousand words] is simple, but it has a major flaw: it does not take into account the possibility that the number of verb phrases per 100,000 words may not be stable diachronically. Nor does it guarantee that the *opportunity* for a progressive to be used is uniform, i.e. as compared to a baseline of progressives plus alternative variants. (2010: 154f.)

In a later paper (Aarts et al. 2013), they compare three different ways of measuring progressive frequency in DCPSE: rate per million words, rate as a proportion of all verb phrases, and rate as a proportion of the number of progressivisable verb phrases (as proposed in their 2010 paper). Interestingly, though, irrespective of the measure, the progressive's frequency increase between LLC and ICE-GB always lies around 20% (2013: 18ff.). Thus, they concede that "[i]n the case of the progressive, our three baselines turn out to be closely aligned over time" (2013: 22).

In the light of these results, Smith and Leech (2013: 73) argue that the difference between simple normalisation on the one hand and measures such as the S-coefficient on the other hand should not be exaggerated. Thus, in their study of progressive use in 20th-century written English, they opt for normalisation per million words.

Furthermore, it has to be noted that the identification of progressivisable verb phrases is by no means straightforward – neither theoretically nor practically. Without discussing

[11] Also called the M-coefficient (cf. chapter 3).

this issue in further detail, it should be clear that simple normalisation is much easier and time-effective.

Based on these considerations, it was decided to measure the progressive's use and development per hundred thousands words (chapter 5, DCPSE & FCSE data) and per million words (chapter 6, BNC94 & BNC14 data), ensuring time-effective analysis and comparability to the majority of previous studies.

The second question posed at the beginning of this section asked how individual variables of progressive use can best be measured to explain the progressive's frequency development. These variables are, for example, verb class or morphosyntactic context. The present study uses two simple frequency metrics that highlight slightly different things.

First, the frequency of an individual variable level, for example the number of progressives with a certain verb class, can be calculated per hundred thousands word – exactly the same way the overall progressive frequencies are calculated. This text-linguistic measure can be regarded as a rough proxy for the variable's entrenchment in the language system – if entrenchment is understood in a broad sense meaning anchoring or routinisation. Let us assume, for instance, that the frequency of the progressive with activity/event verbs has increased by 100 tokens per hundred thousand words between LLC and FCSE. The present study would take this result as an indication that the use of progressives with activity/event verbs has become more entrenched/routinised in the language system, i.e. the community grammar.

Chapter 2.1.1 introduced the concept of entrenchment and argued that it comprises a number of different aspects; most importantly strength of representation, unit status, and degree of schematization. While it was stated that token frequency has an effect on the strength of representation, it is also clear that other aspects of the concept of entrenchment are not simply a function of it and require more sophisticated operationalisation. Notwithstanding these differences between token frequency and entrenchment, De Smet uses the former as a proxy to the latter, arguing that "simple discourse frequency has had undeniable explanatory success in diachronic research" (2017: 77) (see also Hilpert 2017: 52). While Stefanowitsch and Flach generally regard this as a too narrow definition of entrenchment, they nonetheless acknowledge that "there is a correlation between psycholinguistic measures of entrenchment and text frequency in general" (2017: 122). Thus, it seems justified to propose a connection between the normalised corpus frequencies of individual progressive variables and their degree of entrenchment: Those progressive realisations that are very frequent in the community grammar (i.e. the corpus) are clearly more routinised and – most likely – have a stronger mental representation than less frequent uses.

A second way to quantify the frequency of different variables is to measure them relative to their "competitors" – adopting a variationist perspective. Taking the example of progressives with activity/event verbs again, one cannot only quantify their increase per hundred thousand words but also their frequency relative to all other verb classes used with the progressive. This offers a different perspective on the perceived frequency development: Assuming that half of all progressives in LLC occur with an activity/event verb, an absolute increase by 100 tokens per hundred thousand words does not necessarily

translate into a significant increase in the share of activity/event verbs occurring with the progressive. This is due to the fact that other verb classes can exhibit an absolute increase at the same time. In this case, progressives with activity/event verbs would have become more entrenched (in the sense of having strengthened their status in the community grammar), but they would not have increased their share among all progressives – i.e. would not have become more typical of progressive use. This is exactly how the present study interprets changes in relative variable frequencies: as changes in typicality of the respective variable in the progressive's constructional network. Each variable represents a realisation/variant of the progressive construction. If the relative share of variables changes over time – i.e. if one verb class increases its share among progressive uses while another verb class loses ground – the cognitive representation of the progressive's constructional network can be expected to change as well (cf. Hilpert 2017: 54ff.).

This view is closely linked to Hilpert's concept of constructional change (cf. chapter 2.2.2), which places particular importance on relative frequencies: "In the study of constructional change, it is specifically the relative frequency of structural and functional variants of a construction that merits consideration" (2013a: 12).

Apart from these easy-to-calculate frequency metrics, corpus-linguistic methodology offers a broad range of more complex tools. While these certainly have a lot to offer, they are, however, less intuitive and harder to understand. Two of these tools, Collostructional Analysis and Hierarchical Configural Frequency Analysis, will be introduced below.

4.4.2. Mosaic plots

Mosaic plots can be used to visualise and analyse frequency distributions relative to each other. In their easiest form, they are based on a 2x2 contingency table. Figure 4.1 is an example of such a 2-way mosaic plot. The data is entirely made-up and depicts three different levels of linguistic interest among the inhabitants of three German cities.

The overall square size corresponds to the overall number of observations. Each tile depicts the observed frequency of one variable combination. Friendly and Meyer explain it as follows:

> The mosaic display [...] is like a grouped barchart, where the heights (or widths) of the bars show the relative frequencies of one variable, and the widths (heights) of the sections in each bar show the conditional frequencies of the second variable, given the first. This gives an area-proportional visualization of the frequencies composed of tiles corresponding to the cells created by successive vertical and horizontal splits of rectangle, representing the total frequency in the table. (2016: 162)

With regard to the example plot, this means that there are many more people with high linguistic interest in Freiburg than in Heidelberg. Berlin lies in between the two. On the other hand, medium linguistic interest is particularly pronounced in Heildeberg. Low interest is most frequently observed in Berlin. Thus, it should already become apparent how mosaic plots can be used to compare frequencies relative to each other. However, if only the sizes of the tiles are considered, fine-grained statements about the different

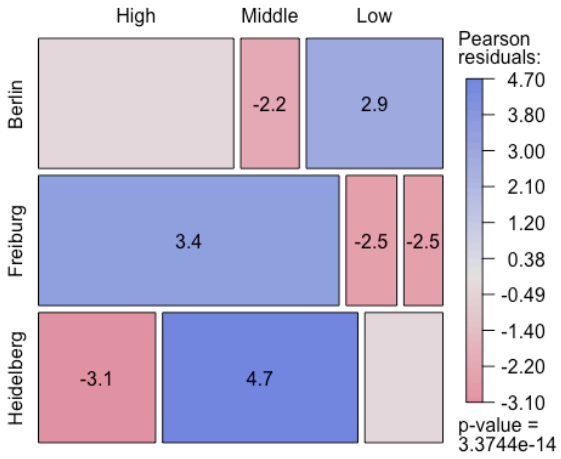

Figure 4.1.: Hypothetical level of linguistic interest in different German cities.

frequencies can hardly be made. This is where the colouring comes in: While the whole data set has a p-value (bottom right) indicating whether there is a significant correlation between the level of linguistic interest and city, the colouring of an individual tile assesses the tile's individual contribution. It indicates whether its observed frequency lies above or below its expected frequency. This deviance is expressed by means of standardised Pearson Residuals (cf. Friendly and Meyer 2016: 304,481). If a residual is positive, the observed frequency of a tile lies above the expected one and the tile is coloured blue (the darker the blue, the more positive the residual). In case of a negative residual, the observed frequency lies below the expected one and the tile is coloured in red (the darker the red, the more negative the residual). This way, the shading levels help us interpreting the plot: In addition to the observed frequencies, represented by the tile size, we can also make statements about the statistical over- or underrepresentation of each tile (i.e. variable combination).

If a residual lies above +2 or below –2, its exact magnitude is indicated in the plot. This is because +/–2 can be seen as a threshold value, indicating that a tile's residual is individually significant at approximately the 0.05 level. Similarly, if a residual lies above +4 or below –4, a significance level of 0.001 can be assumed (Friendly and Meyer 2016: 166). With regard to the example plot, this means that low linguistic interest is not only more frequently observed in Berlin than in Freiburg and Heidelberg, but that it is also significantly more frequent, i.e overrepresented, in Berlin than would be expected in the assumed model of independence. In Freiburg, on the other hand, low interest is significantly less frequent than expected. In Heidelberg, as the very light red shading indicates, low interest is only slightly less frequent than expected (the negative residual is far from reaching the threshold of –2).

Thus, mosaic plots constitute a powerful visual display of frequency distributions relative to each other, assessing whether there is a significant correlation between the

depicted variables, and indicating the status/contribution of individual variable combinations.

4.4.3. Collostructional Analysis

Collostructional analysis (CA) refers to a family of quantitative corpus-linguistic methods that have been developed to measure the strength of association between words and constructions (cf. Stefanowitsch and Gries 2003; Stefanowitsch 2013). It can be seen as an extension of existing collocation-based methods and "its name is a blend of the words *collocation* and *construction*" (Stefanowitsch 2013: 290). In its most basic form, *simple collexeme analysis*, it analyses the frequency of a single construction and compares the frequency of its collocates against the total frequency of these collocates in the corpus.

As psycholinguistic evidence shows, speakers' construction-specific verbal preferences can more faithfully be predicted on the basis of collostructional analysis (i.e. collexemes) than on the basis of simple frequency counts (Gries et al. 2005 & 2010).[12]

In the present work, *distinctive collexeme analysis* (DCA) (Gries and Stefanowitsch 2004) will be used to analyse the progressive's verbal preferences in the different subcorpora. Originally, DCA has been designed to investigate "pairs of semantically similar grammatical constructions and the lexemes that occur in them" (ibid.). It identifies lexemes that exhibit a marked preference for one of two constructions, revealing subtle distributional differences between the members of the pair (ibid.).

Using ICE-GB, Gries and Stefanowitsch (2004: 113ff.) compared the verbal preferences of the so-called *will*-future to those of the so-called *going to*-future. By means of DCA, they identified two sets of distinctive collexemes – one for *will* and one for *be going to*. These sets are made up by verbs that distinguish best between the two constructions, i.e. that highlight differences and hide commonalities. Analysing these two sets qualitatively, the authors were able to approximate the semantic differences between *will* and *be going to*. In a nutshell, they found that *be going to* encodes more dynamic and more specific actions and events (e.g. *say, do, use*) than *will* (e.g. *find, receive, hold*) (2004: 114f.).

DCA cannot only be applied to pairs of semantically similar constructions, but can also be used to investigate the use of one and the same construction in different varieties or periods. The former has been done by Wulff et al. (2007), who investigated the use of the *into*-causative in British and American English. A diachronic application of DCA has been proposed by Hilpert (2006), using it to study meaning change in grammatical constructions. In the case of such a diachronic application, it is in most cases not enough to compare the use of one construction in only two periods. If trends are to be revealed, several periods need to be compared. To this end, a specific version of DCA, *multiple distinctive collexeme analysis* (MDCA), can be used, which enables the comparison of more than two constructions (e.g. Gilquin 2006) or of one construction in different periods (e.g. Hilpert 2012).

[12] The statistical basis of collostructional analysis and its supposed superiority over less complex frequency metrics has been a matter of considerable debate. See, for example, Bybee (2010: 97ff.), Schmid and Küchenhoff (2013) and Gries (2015a).

MDCA can be computed using the coll.analysis3.5 script by Gries (2014).[13] Assuming that we want to compute the distinctive collexemes of construction c in three different periods, i.e. c_1, c_2 and c_3, we need as input the construction's overall frequency in each period as well as the total frequency of each lexeme l_i in each period. The overall number of constructions in the corpus is not required (cf. Stefanowitsch 2013: 297). Table 4.3 gives an overview of the frequency information needed for a DCA analysing two constructions, c_1 and c_2. For our MDCA example with the constructions c_1, c_2 and c_3, we would simply need to add another row with the respective frequency information for c_3 to the table.

	Word l_i of Class L	Other Words of Class L	Total
Construction c_1 of Class C	Freq. of $L(l_i)$ in $C(c_1)$	Freq. of $L(\neg l_i)$ in $C(c_1)$	Total freq. of $C(c_1)$
Construction c_2 of Class C	Freq. of $L(l_i)$ in $C(c_2)$	Freq. of $L(\neg l_i)$ in $C(c_2)$	Total freq. of $C(c_2)$
Total	Total freq. of $L(l_i)$ in $C(c_1,c_2)$	Total freq. of $L(\neg l_i)$ in $C(c_1,c_2)$	Total freq. of $C(c_1,c_2)$

Table 4.3.: Frequency information required for distinctive collexeme analysis (taken from Stefanowitsch 2013: 297).

As Wulff et al. (2009: 360) explain, Gries' coll.analysis3.5 script

> uses an exact binomial test to quantify the association strength between the verbs and their [...] [constructional realizations]. More precisely, it provides a *p*-value for each verb with each [...] [construction] and log-transforms it such that highly positive and highly negative values indicate a large degree of attraction and repulsion, respectively, and 0 indicates random co-occurrence. An (absolute) *p*-log value that is equal to or higher than 1.3 corresponds to a probability of error of 5% or less. (2009: 360)

Thus, if lexeme l_i had a *p*-log value of 1.8 with construction c_3 (i.e. with construction c in period 3), it could be considered a distinctive collexeme that is significantly attracted to c_3. For the present study, this would mean that a certain verb$_i$ is significantly attracted to the progressive construction (c) in FCSE (period 3).

It is important to realise that both DCA and MDCA do not simply "reward" high frequencies. Among the distinctive collexemes one will always find lexemes with high and low co-occurrence frequencies. This is because the procedure is aimed at detecting fine semantic differences. These would not become apparent if one analysed a single construction in isolation and ranked its co-occurring lexemes according to frequency (cf. Hilpert 2012: 140). As Hilpert states, "[d]istinctive collexeme analysis allows the researcher to abstract away from elements that are frequent in both constructions. Instead, it determines whether there are asymmetries in the relative frequencies of the co-occurring lexical verbs" (ibid.).

In chapter 5.2.2, MDCA will be used to investigate the progressive's recent semantic development on individual verb level. The construction's verbal choices in LLC, ICE-GB and FCSE will be compared on the basis of the distinctive collexemes. This way,

[13]The script is available from its author.

it will be possible to reveal specific lexical-semantic changes that have occurred in the construction's recent history.

4.4.4. Hierarchical Configural Frequency Analysis

Another statistical method that can be used as a typicality measure is *Hierarchical Configural Frequency Analysis* (HCFA) (Von Eye 1990; Von Eye 2002; Gries 2009: 240ff.). Other than Collostructional Analysis, HCFA was not originally developed as a corpus linguistic tool and is, for example, used in psychological or medical research (Hilpert 2013a: 56). However, HCFA constitutes a statistical procedure that can very fruitfully be applied to corpus data. Its main aim is to detect combinations of categorical variables that are significantly more or less frequent than expected in a data set. Since the analysis is not limited to detecting low-level combinations of just two variables, but can actually find high-level interactions of, for example, four or five variables, it can be used to search the data for specific subschemas of a linguistic construction in a bottom-up fashion.[14] If a particular variable combination is significantly more frequent than expected, it is classified as a *type*. In case of statistically significant underrepresentation, one speaks of an *anti-type*.

But how exactly can such a procedure be of use in the analysis of the progressive's recent frequency development? Let us assume the following hypothetical scenario: The analysis of the corpus data reveals that one morphosyntactic realisation of progressive use – for example progressives occurring in connection with modal auxiliaries – has increased in frequency. Furthermore, the analysis shows that one verb class – e.g. stance verbs – has also become considerably more frequent in recent decades. While the morphosyntactic change has occurred in all genres, the increase in stance verbs has only occurred in Face-to-face Conversation. Using HCFA, it would be possible to test how the four variables – morphosyntactic context, verb class, genre, and subcorpus – interact. Theoretically, it could be the case that the increase in stance verbs in Face-to-face Conversation did not happen independently of the increase in modal progressives, resulting in significant overrepresentation of the variable combination *modal progressive + stance verb + Face-to-face Conversation + FCSE* – making this combination a *type*.

While this example is strictly hypothetical and, most likely, will have no correspondence in the data, it highlights the decisive advantage of HCFA in the analysis of categorical data: Instead of just analysing the development of different variables independently of each other, HCFA exposes possible interactions of several categorical variables, enabling the researcher to reveal the existence and the behaviour of very specific realisations of a linguistic construction. As Hilpert (2013a: 65) argues, these specific realisations/variable combinations can be regarded as subschemas of a higher-level construction, meaning that HCFA allows the researcher to investigate a constructional network on a fine-grained level. If, as in our example, time is included as a variable (using data from different

[14] To be precise, the procedure is not entirely bottom-up since it depends on the variables that are fed into the analysis. However, within the boundaries of the available variables, the methods functions in a bottom-up fashion since it does not rely on any preconceived assumptions about the variables' possible combinatorial preferences.

subcorpora), it becomes possible to study changes in this constructional network by revealing how specific variable combinations interact over time.

HCFA can be conducted using an R script written by Gries (2004), which takes as input the raw data of all tokens and their variable levels. Statistically, the analysis is an extension of the chi-square test, which is applied to all possible variable combinations (in addition to a global test of the whole table). The exact procedure is concisely explained by Hilpert:

> HCFA assesses for each cell in the table whether its contribution to the overall chi-squared value exceeds a certain threshold. Assessing each individual cell after the fact means that what is tested is no longer a single null hypothesis (H0: observed frequencies = expected frequencies) but in fact a series of null hypotheses, each of which holds that for a given cell in the table, the observed frequencies will roughly equal the expected frequencies. Whenever multiple hypotheses as tested, the required significance levels have to be adjusted downward, that is, the usual $p<0.05$ needs to be made smaller. This can be done through a procedure such as the Bonferroni correction [...], which determines the level of p by dividing the desired value (such as 0.05) by the number of cells in the table. (2013a: 62f.)

In the present study, however, it was not the Bonferroni correction that was used for downward adjustment of p-values but the – just as conservative – Holm adjustment, which can detect more significant configurations (Gries 2009: 249).

The analysis hierarchically assesses all possible distributions. Again, Hilpert explains: "In addition to cross-tabulations of all [...] variables, the HCFA further provides significance tests for the univariate distributions and all lower-level configurations. The analysis targets the full hierarchy of configurations, from the simple to the complex" (2013a: 64). For each variable and for all possible combinations, the output indicates the observed and the expected frequency and provides a $>$ or $<$ sign depending on whether a combination is more or less frequent than expected. In addition to the exact contribution to the overall chi-square value, the table lists the Holm-adjusted p-value and an indication of statistical significance. Furthermore, an effect size Q is returned, which indicates the configuration strength of the variable combinations. As already mentioned, those combinations that are significantly over- or underrepresented are called types or anti-types (being typical or atypical of the data set).

It could be argued that it is only at this point that the real work of the researcher starts: All types and anti-types have to be interpreted and the actual tokens need to be thoroughly studied. Thus, a purely quantitative procedure is followed by a qualitative analysis that has to make sense of the numerical output. In fact, this is very similar to MDCA, where a list of distinctive collexemes alone does not constitute an insightful analysis.

4.4.5. Logistic regression

I would like to conclude this chapter by introducing a statistical procedure that has widely been used in corpus linguistics: binary logistic regression. It models the outcome of a binary dependent variable – such as *will* vs. *going to* in future marker choice – based on one or more independent variables that can be categorical and/or continuous. Thus, it differs from MDCA and HCFA in that it is not exploratory but focused on one specific linguistic choice between two alternatives. Examples of influential corpus linguistic studies that have applied logistic regression and that are conceptually relevant for the present work are Szmrecsanyi (2005 & 2006) and Hilpert (2013a). Good introductions are Gries (2013b: 293ff.) and Baayen (2008: 195ff.). A very thorough treatment of the topic – however, not specifically aimed at linguists – is provided in Field et al. (2012, chapter 8).

In the present work, logistic regression will be used to model speakers' aspectual choices in the use of lexically specific patterns (chapter 6). Based on corpus, verb, aspectual encoding of the previous realisation, and recency, progressive or simple aspect choice of a pattern will be predicted. Since regression models cannot only handle main effects of individual variables but also interactions between them, it will be possible to analyse how different verbal realisations of a pattern have developed over time. While progressive encoding of a specific pattern can have become more frequent with one verb (e.g. *I'm not saying*), it may have decreased with other verbs (e.g. *I'm not going*) (relative to the number of corresponding simple aspect encodings; i.e. *I don't say/go*). For a semi-fixed pattern such as *I'm not -ing*, logistic regression will reveal exactly which verbal realisations are most fixed in their progressive encoding (meaning that they are rarely used in their corresponding simple aspect encoding) and to what degree this has changed in the course of the past decades.

As has already been stated, the regression models will take into account two further variables apart from corpus and verb: aspectual encoding of the previous realisation, and recency. The first variable indicates whether the previous realisation of a pattern was realised with the same aspectual encoding as the current realisation (i.e. progressive > progressive, or simple > simple). As Szmrecsanyi (2005 & 2006) and Gries (2005 & 2011) have shown, syntactic priming effects (cf. Bock 1986; Bock et al. 2007) are pervasive in dialogic corpus data: Speakers tend to re-use the syntactic structure of a recently produced or comprehended utterance, which means that we can expect them to have a tendency to maintain the aspectual encoding of a pattern (e.g. *I don't think > I don't suggest* vs. *I'm not saying > I'm not suggesting*) (cf. also chapter 2.1.1). Since such priming effects account for a considerable degree of variation in spoken dialogue (Szmrecsanyi 2006), it is only logical to include them in the regression model to achieve a good fit to the data.[15] The second variable, recency (operationalised as distance in

[15] In order to take into account the fact that production-to-production priming is stronger than comprehension-to-production priming (cf. Branigan et al. 1995; Branigan et al. 2000), the model will also consider whether the previous realisation of a pattern was produced by the same speaker or his/her interlocutor.

tokens to the previous realisation[16]), is expected to interact with the priming effect: As distance increases, the strength of a prime decreases. That such a relationship between priming and recency can be made visible in spoken corpus data has for example been shown by Szmrecsanyi (2005).

Logistic regression can be implemented in R using the *glm* function of the Stats package (R Core Team 2017), where *glm* stands for *generalized linear model*.[17] In order to be able to use a linear model to predict the outcome of a binary dependent variable, the latter

> is transformed with a so-called link function, which transforms the predicted range of values of a linear model (-∞ to +∞) to a range more appropriate for the dependent variable. For binary logistic regression, [...] the inverse logit transformation [...] transforms values from the range of -∞ to +∞ into values ranging from 0 to 1, which can then be interpreted as probabilities of a predicted event. (Gries 2013b: 293)

Thereby, the problem of violating the assumption of linearity is overcome (Field et al. 2012: 315).

Very much the same as in linear regression, each predictor variable in the logistic regression equation gets its own coefficient (b) (ibid.). As Field et al. explain, this coefficient

> represents the change in the *logit* of the outcome variable associated with a one-unit change in the predictor variable. [...] The crucial statistic is the z-statistic, which has a normal distribution and tells us whether the b coefficient for that predictor is significantly different from zero. If the coefficient is significantly different from zero then we can assume that the predictor is making a significant contribution to the prediction of the outcome (Y). (2012: 332f.)

One way of interpreting the regression results for the different coefficients is odds ratios, which are the exponentials of b and indicate the change in odds resulting from a unit change in the predictor. They range from 0 to ∞, and if their value is greater than 1, "then it indicates that as the predictor increases, the odds for the outcome occurring increase. Conversely, a value less then 1 indicates that as the predictor increases, the odds of the outcome occurring decrease" (Field et al. 2012: 319f).[18]

[16] More precisely, it is the natural logarithm *ln* of the distance in tokens that was used as predictor variable. As Gries (2005: 374) explains, "[d]istance [...] [is] entered into a linear model, but [...] the relation between the distance between prime and target on the one hand and the strength of the priming effect on the other hand need not be linear. In fact, there is evidence that this relation is logarithmic." (see also Gries 2007: 277).

[17] A statistically more sophisticated version would have been offered by the *glmer* function, which is capable of accounting for random, i.e. non-fixed, effects such as speaker- or text-ID. In doing so, the variability inherent in the data can better be accounted for (cf. Gries 2011 & 2015b). Unfortunately, though, it was not possible to construct converging mixed-effects models based on the data used in chapter 6.

[18] See Gries (2013b: 300, Figure 72) for an instructive visualisation of three alternative ways of interpreting binary logistic regression results.

Apart from the coefficients, a number of model statistics can be calculated. One is the model chi-square statistic χ^2, which tests "whether or not all of the variables included in the model significantly contribute to explaining the variance in the dependent variable" (Szmrecsanyi 2006: 58). Furthermore, it is possible to obtain a pseudo R^2 statistic for logistic regression (*Nagelkerke's* R^2), which is an indication of the model's overall explanatory power. It ranges form 0 to 1 and indicates how much of the variance in the dependent variable is explained by the independent variables included in the model. Values above 0.2 are considered as acceptable, values above 0.4 as good, and values above 0.5 as very good (cf. Szmrecsanyi 2006: 58; Field et al. 2012: 316ff.).

One potential problem that can occur in binary logistic regression – just as in any other regression model – is multicollinearity. It arises if two or more independent variables show a strong correlation, leading to unreliable estimates of the b coefficients (cf. Field et al. 2012: 274ff.; Szmrecsanyi 2006: 54). Several collinearity diagnostics can be calculated in R, one of which is the *variance inflation factor* (VIF). As Field et al. explain, "the VIF indicates whether a predictor has a strong linear relationship with the other predictor(s)" (2012: 276). Values below 10 are acceptable, while values above 10 indicate multicollinearity (cf. also Szmrecsanyi 2006: 215).[19] All independent variables of the regression models calculated in chapter 6 were checked for their respective VIFs.

[19] It has to be noted, however, that there are cases of multicollinearity that are no cause for concern. One such case concerns high VIFs that are caused by the inclusion of interaction terms in the model. As Allison (2012) explains, "if your model has x, z, and xz, both x and z are likely to be highly correlated with their product. This is not something to be concerned about [...]." According to Frost (2017), this type of collinearity is structural, meaning that it can be seen as a byproduct of the model rather than being present in the data itself.

5. The progressive's development and use in spoken 20th and 21st-century English

This chapter is structured as follows: Section 5.1 addresses Hypothesis 1, reporting the construction's frequency development in different genres. The following two sections are concerned with Hypothesis 2, analysing the observed frequency developments in detail. Section 5.2 analyses the development of two individual variables – morphosyntactic context and verb. Section 5.3 focuses on the interplay of variable combinations, reporting the results of the Hierarchical Configural Frequency Analysis.

5.1. Frequency development in different genres

Based on a comprehensive summary of previous research on the progressive's frequency development, the following hypothesis was formulated in chapter 3.5.2:

Hypothesis 1 The frequency increase of the progressive is still under way in the 21st century, and it is proceeding from different baselines at different speeds in different genres.

Before turning to the actual numbers, it has to be noted that Broadcast Interviews and Broadcast Discussions will be treated as one genre. The same is true for Parliamentary Language and Prepared Speech. This is due to the fact that especially Parliamentary Language and Broadcast Interviews have very small word numbers in DCPSE. Consequently, the number of progressive tokens obtained for these genres was low and, arguably, not very reliable. Thus, it was decided to conflate these genres with closely related ones. For Broadcast Interviews the closest match clearly were Broadcast Discussions. The transcripts of both genres stem from TV or radio shows in which at least one host talks to one or several studio guests. While Broadcast Discussions seem more involved and always entail more than two interlocutors, Broadcast Interviews mostly have only two speakers. In general, however, the two genres cannot be sharply divided, since also an interview situation between one interviewer and one interviewee can constitute an involved discussion. Thus, the conflation of the two genres should be fairly unproblematic. Similarly, Parliamentary Language and Prepared Speech share a lot of similarities: Probably most important is the fact that a lot of language used in parliamentary debates is scripted, as are the texts in Prepared Speech. Parliamentary speeches are normally read out and also questions are often pre-phrased and read from paper. Furthermore, both genres contain a lot of expository passages and are heavily influenced by the norms of written language. Based on these considerations, it seems

justified to treat them as one genre. With regard to one aspect, however, the two genres differ: While the great majority of texts in Prepared Speech have only one speaker (with a few exceptions in ICE-GB), transcripts of parliamentary debates are characterised by a very high number of different speakers (the difference between the two genres is much more pronounced than the one between Broadcast Interviews and Discussions). This goes back to the nature of parliamentary debates: The speaker who currently has the floor has to continuously answer questions from other members of parliament. Thus, the parliamentary transcripts in LLC, ICE-GB, and FCSE have an interactive character that is not present in the transcripts of Prepared Speech. Consequently, one might object to treating the two genres as one. The reasons for doing so nonetheless are the following: First, the interactive character of parliamentary debates is a highly routinised one, frequently entailing scripted questions and answers, differing considerably from conversational exchanges/discussions. Second, as explained above, except for the difference in speaker number there are several commonalities between the two genres. Third, progressive frequency does not seem to be closely correlated with speaker number. For example, progressive frequencies in DCPSE parliamentary debates (high number of speakers) are low, while they are high in FCSE Broadcast Interviews (low number of speakers). Fourth, conflation of the two genres results in a sounder data base for the DCPSE data, enabling more robust interpretations.

The corpus frequencies – absolute and/or normalised per hundred thousand words – are visualised in Figure 5.1 and summarised in Table 5.1.[1]

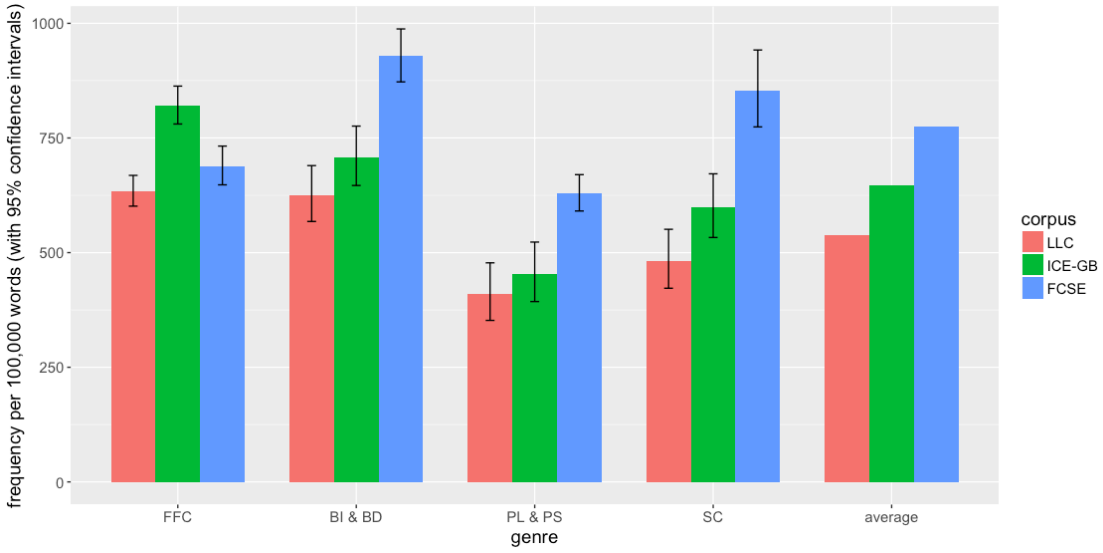

Figure 5.1.: Progressive frequency by corpus and genre. FFC = Informal Face-to-face Conversation, BI & BD = Broadcast Interviews and Discussions, PL & PS = Parliamentary Language and Prepared Speech, SC = Spontaneous Commentary.

[1]The numbers for each individual genre are provided in the appendix, Table A.2.

	LLC			ICE-GB			FCSE			Difference LLC - ICE-GB		Difference ICE-GB - FCSE		Difference LLC - FCSE	
Genre	Raw freq	word no.	**freq htw**	Raw freq	word no.	**freq htw**	Raw freq	word no.	**freq htw**	sign. level	change in %	sign. level	change in %	sign. level	**change in %**
FFC	1384	218307	**634**	1523	185537	**821**	1031	149690	**689**	***	+29.5%	***	–16.1%	*	**+8.7%**
BI & BD	414	66136	**626**	468	66067	**708**	1002	107935	**928**	*n.s.*	+13.1%	***	+31.1%	***	**+48.2%**
PL & PS	171	41676	**410**	195	42982	**454**	974	154813	**629**	*n.s.*	+10.7%	***	+38.5%	***	**+53.4%**
SC	224	46426	**482**	293	48955	**599**	406	47540	**854**	*	+24.3%	***	+42.6%	***	**+77.2%**
Average			538			646			775		+20.1%		+20.0%		+44.1%

Table 5.1.: The progressive's frequency development in DCPSE and FCSE (Chi-square test for statistical significance: *** p<0.001, ** p<0.01, * p<0.05) htw = hundred thousand words, FFC = Informal Face-to-face Conversation, BI & BD = Broadcast Interviews and Discussions, PL & PS = Parliamentary Language and Prepared Speech, SC = Spontaneous Commentary.

The results reveal an overall increase of the progressive between LLC and FCSE in all genres. The construction is more frequent today than it was in mid-twentieth-century spoken English. This overall increase is statistically significant for all genres. The development follows a consistent trend in Broadcast Interviews & Discussions, Parliamentary Language & Prepared Speech, and Spontaneous Commentary, meaning that an increase can be observed both between LLC and ICE-GB, and ICE-GB and FCSE. For Face-to-face Conversation, however, no such trend is visible: While the progressive significantly increases in frequency between LLC and ICE-GB, the numbers significantly decrease between ICE-GB and FCSE. Moreover, Face-to-face Conversation loses its status as genre with highest progressive frequencies, which it held in LLC and ICE-GB. In FCSE, the progressive is most frequent in Broadcast Interviews & Discussions, and also Spontaneous Commentary exhibits a higher use of the progressive than Face-to-face Conversation. In all corpora, the progressive is least frequent in Parliamentary Language & Prepared Speech. Averaging over the normalised genre frequencies reveals a steady 20% increase between LLC and ICE-GB, and ICE-GB and FCSE. Thus, the progressive's frequency increase is clearly still under way in 21st-century spoken British English.

The most pronounced increase between LLC and FCSE can be observed for Spontaneous Commentary (+77.2%), followed by Parliamentary Language & Prepared Speech (+53.4.%), Broadcast Interviews & Discussions (+48.2%), and Informal Face-to-face Conversation (+8.7%). The increase in Spontaneous Commentary, Parliamentary Language & Prepared Speech, and Broadcast Interviews & Discussions has been considerably more pronounced between the 1990s and today than between the mid-20th century and the 1990s, meaning that it has accelerated in recent decades. All of these genres can be considered to be more formal than Face-to-face Conversation, i.e. more influenced by written norms. Broadcast genres – i.e. interviews, discussions, and live commentary – are regulated by editorial guidelines of the broadcast company, for example controlling the use of so-called "strong" language. There are, however, also a multitude of implicit rules or expectations how to communicate "properly" on the radio or on TV, meaning that language use in these genres can be expected to deviate from authentic, informal conversation in several respects. The same is true for parliamentary debates, lectures, or public speeches. However, as has been observed for written English, the second half of the 20th century has seen developments such as colloquialization and democratization of discourse phenomena (cf. Mair 2006b: 183ff.). In the case of colloquialization, the written norm has moved closer to spoken usage. This trend can also be expected to apply to conservative spoken genres, and the observed increase of the progressive in broadcast genres, parliamentary debates, lectures, sermons, and public speeches can be interpreted as a result of it. Since the progressive has always been more frequent in spoken conversation than in written language, it constitutes a feature of conversational style. Its rapid spread in written 20th-century English (cf. chapter 3.3) and the figures presented here indicate that in the course of the colloquialization of formal written and spoken styles the construction has continually diffused into the respective genres and is undergoing a frequency increase to the present day.

5.1.1. Accounting for the development in Face-to-face Conversation

Let us now have a closer look at the frequency development in Face-to-face Conversation. Unlike in all other genres, progressive use is most frequent in ICE-GB, showing a decrease between the 1990s and today. How can this development be explained? First, the numbers could reflect a saturation point of progressive use in conversation towards the end of the 20th century. This explanation, however, would not explain the subsequent decrease in the FCSE data, for which a frequency plateau would be expected instead. Especially against the background of the still increasing progressive frequencies in all other genres this recent decrease in Face-to-face Conversation cannot straightforwardly be explained.

It seems conceivable that Face-to-face Conversation is a much more variable and less unified genre than is often tacitly assumed: The data used for the different subcorpora must vary along certain lines that are not adequately captured by the genre classification as Face-to-face Conversation. For example, narrative passages are more frequent in the conversational data of FCSE than in ICE-GB conversations.[2]

A comparison of the results to progressive frequencies in the much larger BNC-DS (BNC94)[3] and Spoken BNC2014 (BNC14)[4] supports this claim.

	freq per htw[1]	word number
BNC94	~ 644 (~ 32,314)	5,014,655
BNC14	~ 759 (~ 86,676)	11,422,617
	***+17.86%	

Table 5.2.: Approximated frequency development (per hundred thousand words) of the progressive in BNC-DS (conversation) (BNC94) and Spoken BNC2014 (BNC14). [1]Numbers in brackets indicate approximated raw frequencies.

As can be seen in Table 5.2 and Figure 5.2, the estimated level of progressive use in BNC14 lies at 759 tokens/htw.[5] Thus, it is higher than the one reported for FCSE

[2]A narrativity analysis was performed for a sample of the ICE-GB and FCSE Face-to-face Conversation texts. While it did reveal a higher level of narrative passages (i.e. narrations, stories, accounts) in the FCSE texts (see Figure A.3, appendix), this was not linked to less frequent use of the progressive. On the contrary, narrative passages are characterised by increased progressive use (see Figure A.4, appendix). Thus, differences in narrativity fail to explain the less frequent use of the construction in FCSE Face-to-face Conversation.

[3]The conversational data from the demographically sampled (DS) part of BNC dates back to the 1990s (cf. chapter 4.2.3) and can be compared to the results of ICE-GB Face-to-face Conversation.

[4]Spoken BNC2014 contains transcripts of present-day spoken conversation, making it a match for the respective FCSE data.

[5]Since Spoken BNC2014 only became available in late 2017 – at a time when the present project had already been far advanced –, it was not possible to undertake a comprehensive analysis of all progressive tokens in the corpus and to compare the two BNC corpora in a detailed way. Nonetheless, it was decided to make use of BNC94 and BNC14 by approximating progressive frequencies in both of them. The corpora were searched for a form of *be* followed by a present participle, allowing for up to three intervening elements. These intervening elements were further restricted to achieve better

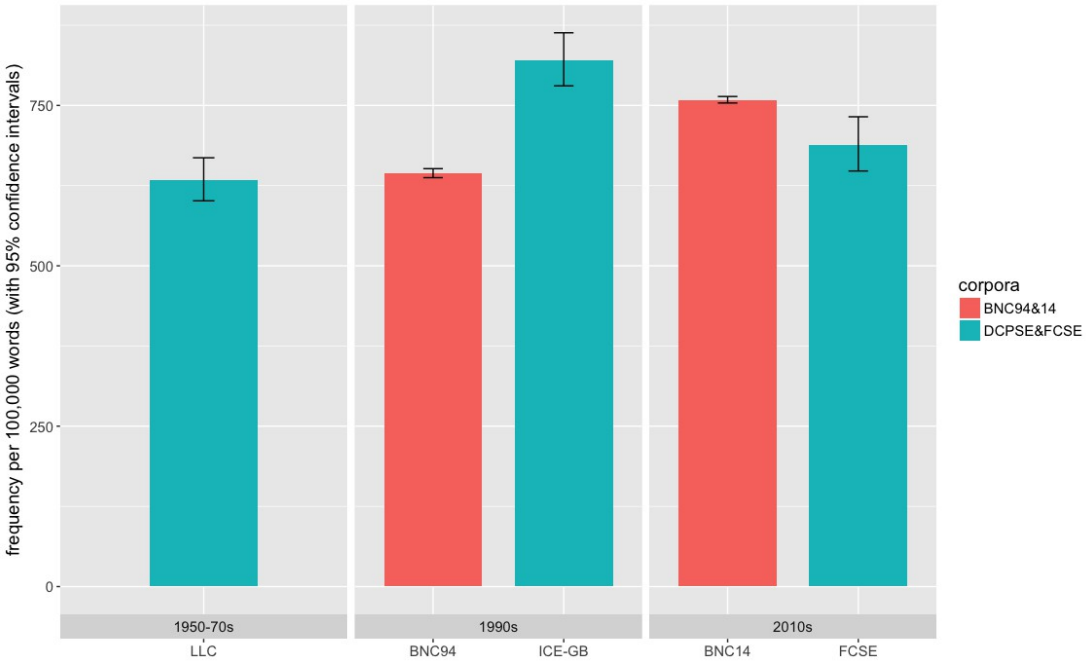

Figure 5.2.: The progressive's frequency development in DCPSE, FCSE, and BNC Face-to-face Conversation (per hundred thousand words). BNC frequencies approximated.

Face-to-face Conversation (689 tokens/htw). However, taking into account the relevant confidence interval, the difference does not appear all too dramatic and seems to reflect the fact that we are dealing with corpora of different size that were compiled in slightly different ways. Much more problematic is the comparison of the BNC94 result (644 tokens/htw) with the level of use in ICE-GB Face-to-face Conversation (821 tokens/htw). The former is much lower than the latter, seriously undermining the claim that we are dealing with comparable kinds of data. Either the approximated level of use for BNC94 is flawed or the progressive is overrepresented in ICE-GB Face-to-face Conversation. Overall, Table 5.2 reports a frequency increase of ca. 18% between the late 1990s and today, which comes from a relatively low level in BNC94 (basically the level reported for LLC Face-to-face Conversation). These results do not support the ones found for

results. For example, they did not allow for a present participle, punctuation marks, prepositions (to exclude false positives such as *he was out running*) or adjectives (to exclude false positives such as *I'm happy walking*). In a second step, a sample of several hundred tokens was extracted both from the BNC94 and the BNC14 results and manually checked for false positives. This way, it was possible to come up with an estimated error rate for both kinds of output, enabling the calculation of the approximate level of progressive use in the two corpora. While these results cannot be regarded as definitive, they are nonetheless likely to provide a relatively valid approximation of the respective progressive frequencies.

DCPSE and FCSE, especially running counter to the level of use in ICE-GB. They support Hypothesis 1, which proposed a continued frequency increase in all genres.

What remains is an insight that can hardly be satisfying for the corpus linguist: Based on two different sets of corpora – DCPSE & FCSE and BNC94 & BNC14 – , which should be comparable in theory, the analysis has arrived at different results. Since both data sets clearly contain truly spoken conversation, this can only mean that conversational data is inherently more variable than is generally accepted (see chapter 7 for a discussion). Consequently, it seems doubtful whether Face-to-face Conversation should really be regarded as a supposedly natural baseline against which changes in other genres can easily be compared.

The remaining part of this chapter will further analyse the progressive output from DCPSE and FCSE, including the data from Face-to-face Conversation, which is regarded as evidence in its own right. However, the connection to the BNC results will be made where necessary.

5.1.2. Comparison to previous results

Let us now compare the present results to previous findings on the progressive's frequency development in 20th-century spoken English, presented in Chapter 3.3.1.

Disregarding genre-specific developments, Aarts et al. (2010 & 2013) arrived at an overall increase of 22.13% between LLC and ICE-GB (i.e. in DCPSE). Even though they searched a non-official version of DCPSE, used a different retrieval procedure, and their overall increase was calculated differently than the average increase presented in Table 5.1[6] (20.10%), the numbers are very similar.

Smith (2005: 76) and Leech et al. (2009: 289) presented genre-specific developments for the second half of the 20th century based on spoken mini-corpora. While Smith reported a slight decrease of progressive frequencies in conversation, assuming that a saturation point had been reached by the early 1960s, Leech et al. observed a major increase between their mini-corpora, arguing that the progressive's frequency development in conversation was still under way in late 20th-century spoken English. The present results from BNC94 and BNC14 support this claim, indicating an ongoing increase in Face-to-face Conversation. Since these findings are backed by a more solid data base than the ones from DCPSE and FCSE, it appears safe to adopt the view that the progressive's frequency increase in Face-to-face Conversation is still under way.

Both Smith (2005) and Leech et al. (2009) reported a significant frequency increase for sports/ceremonial commentary. This positive trend is confirmed by the present study, which observes a significant frequency increase of 24.3% for Spontaneous Commentary between LLC and ICE-GB, continuing in an even more pronounced way between ICE-GB and FCSE (+42.6%). Smith and Leech et al. also presented numbers for Broadcast Interviews and Discussions. For Broadcast Discussions, Smith observed an increase of 19% between his mini-corpora, while Leech et al. report an increase of 49.7%. The

[6] I calculated the average of the normalised frequencies, resulting in equal weighting of each genre. Aarts et al. (2010) arrived at the figure of 22.13% by normalising the overall frequencies of each subcorpus, meaning that the contributions of large genres are more influential than those of small ones.

present study finds an overall increase of 13.1% in Broadcast Interviews and Discussions between LLC and ICE-GB. When studied individually, Broadcast Discussions exhibit an increase of 24% in the same period. Again, the present study supports the positive trends reported in previous studies. Finally, for Broadcast Interviews alone, Leech et al. report a non-significant decrease of 14.8% for late 20th-century spoken English (which is based on very low overall numbers). Similarly, in the present study, progressive frequencies in Broadcast Interviews also show a slight negative trend between LLC and ICE-GB (which is based on low raw frequencies and not statistically significant, though).

If a wider diachronic perspective is adopted, the present results show that the progressive's long-term frequency increase is continuing at a dramatic rate. Within roughly half a decade, the construction's frequency has increased by ca. 50% or more in Broadcast Interviews & Discussions, Parliamentary Language & Prepared Speech, and Spontaneous Commentary. Notwithstanding the observed decrease in Face-to-face Conversation between ICE-GB and FCSE, we still arrive at an overall increase of ca. 9% between LLC and FCSE – which, in the long run, is considerable. Comparing the level of use in LLC to the one in BNC14, this rate increases to ca. 20%.

If we expect the written medium to follow the spoken mode, we can predict a continuation of the progressive's frequency increase in written English. Thus, the much-studied frequency development of the progressive construction can be assumed to keep corpus linguists busy in days to come.

5.2. Analysing the frequency development I: Individual variables

The focus is now put on the development of individual variables of progressive use. In chapter 3.5.2, the following hypothesis was formulated:

Hypothesis 2 The frequency increase of the progressive in the 20th and 21st centuries is based on further expansion of prototypical core uses as well as on structural or functional innovations. Structural and functional innovations are expected to have less impact on the corpus data than the development of the prototypical core uses.

This subchapter looks at two central aspects of progressive use – morphosyntactic realisations and the construction's verbal paradigm – and analyses how the respective variables have developed in DCPSE and FCSE. This way, I expect to shed light on the driving forces behind the overall development and to answer how the constructional network of the progressive has changed over the past decades in the wake of the construction's frequency increase.

5.2.1. Morphosyntactic context

Alongside the progressive's verbal paradigm, its morphosyntactic realisations have received considerable attention in previous research and are regarded a key factor in the

analysis of the construction's frequency development. This section focuses on the different morphosyntactic realisations of the progressive active.[7] Since use of the progressive passive is comparatively rare and shows no significant development, it will not be addressed in detail.[8]

Development per hundred thousand words

As has been explained in chapter 3.3.2, the progressive's frequency increase in 20th-century English has been mainly driven by the most frequent form, the present progressive active. While certain less frequent morphosyntactic variants also showed an upward trend – for example progressives with modal auxiliaries (cf. Smith 2005: 70) –, it was the present progressive active that was identified as the most prototypical form and the most important frequency driver.

Present progressive Figure 5.3 depicts the development of the progressive's different morphosyntactic contexts by genre.[9] The numbers are normalised per hundred thousand words (htw) and displayed with 95% confidence intervals.

Regarding the present progressive,[10] two things can be observed immediately: First, it is the most common progressive use in each subcorpus and genre. Second, in all genres it is more frequent in FCSE than in LLC, meaning that its use has become more frequent in the course of the past 60 years. This increase is statistically significant for all genres.[11]

The progressives in sentences 5.1 to 5.4 are examples of present progressive use in FCSE.

(5.1) *Yeah, both teams* ***are really struggling*** *to find their way into this game.* (FCSE_SC_4_97)

(5.2) *At the moment they* ***'re making*** *the wrong choice.* (FCSE_BI_13_8)

(5.3) *Mr Speaker I have set out the three key immediate challenges we* ***are facing****.* (FCSE_PL_2_56)

(5.4) *I* ***am not joining in*** *because it is wrong.* (FCSE_FFC_3_465)

When comparing the progressive's overall frequency development in the different genres (Figure 5.1) to the observed morphosyntactic changes, it becomes apparent that the development of the present progressive almost mirrors the overall trend – revealing how closely linked the two phenomena actually are. For three genres – Broadcast

[7] The frequency development of the progressive active is very similar to the overall development including passives. It is depicted in Figure A.1 (appendix).

[8] The progressive passive's development is depicted in Figure A.2 (appendix).

[9] A display without present and past progressives is provided in Figure A.5 in the appendix. There, changes in the less frequent morphosyntactic contexts become more obvious.

[10] From now on, if not stated otherwise, the label *present progressive* refers to the present progressive active only.

[11] Cf. Table A.3 for the exact numbers and significance levels.

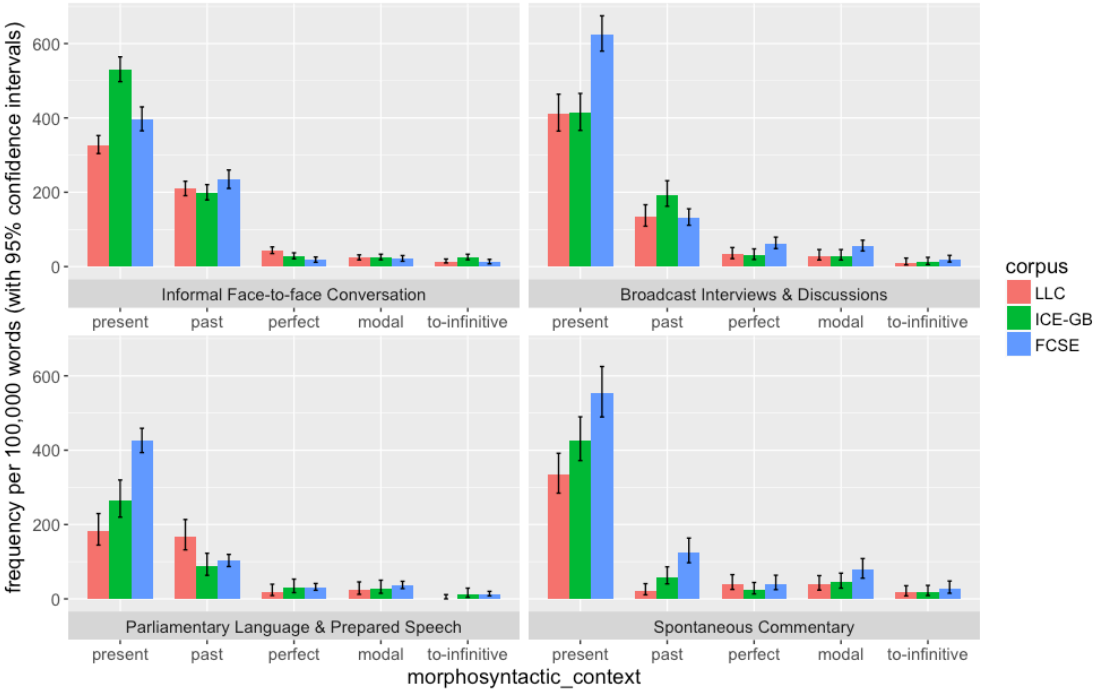

Figure 5.3.: Morphosyntactic context by corpus and genre (active progressives only). pr = present, pa = past, per = present & past perfect, mo = modal, toI = *to*-Infinitive.

Interviews & Discussions, Parliamentary Language & Prepared Speech, and Spontaneous Commentary – a continuous, unbroken frequency increase can be observed (for Broadcast Interviews & Discussions, though, the upward trend between LLC & ICE-GB is extremely small (+0.5%)). In Face-to-face Conversation, the present progressive increases dramatically between LLC & ICE-GB, followed by a significant – however, less pronounced – decrease, also mirroring the overall development.

Adopting a broad perspective and only focussing on the difference between LLC and FCSE, one cannot only state that the present progressive has become significantly more frequent per htw in all genres but also that it is the strongest single contributor to the progressive's overall frequency increase. Measuring how many more present progressive tokens per htw are used today than in the mid-twentieth century, one sees that in Broadcast Interviews & Discussions the present progressive has increased by 214 tokens/htw. In Parliamentary Language & Prepared Speech the change rate is 243 tokens/htw. In Spontaneous Commentary an increase of 219 tokens/htw can be observed. Even in Face-to-face Conversation, where an increase between LLC and ICE-GB is followed by a decrease between ICE-GB and FCSE, the overall increase rate of the present progressive between LLC and FCSE is still as high as 69 tokens/htw. In all genres the present progressive's change rate is higher than the change rate of all the other morphosyntactic realisations. Thus, the present progressive has been the major frequency driver behind

the progressive's overall increase over the past 60 years. These results corroborate previous findings and support Hypothesis 2.

Before turning to the other morphosyntactic realisations, though, it is necessary to focus on the present progressive's development between ICE-GB and FCSE in Face-to-face Conversation. As briefly mentioned above (chapter 5.1.1), FCSE Face-to-face Conversation is characterised by a higher share of narrative passages than ICE-GB Face-to-face Conversation (see Figure A.3). Assuming that narratives are characterised by more frequent use of the past tense, one could argue that the decline in the present progressive is simply a logical consequence of the different narrativity scores. However, if the change rates of the present and the past progressive between ICE-GB and FCSE are compared, it becomes apparent that the increase of the past progressive (+35 tokens/htw) by no means compensates for the decline of the present progressive (–134 tokens/htw). Thus, the increased occurrence of narratives in FCSE Face-to-face Conversation indeed leads to a higher rate of past progressive use, but this increase is much less pronounced than the decrease of the present progressive. Since narrative passages are characterised by more frequent progressive use than non-narrative passages (see Figure A.4), the fact that the present progressive decreases more than the past progressive increases cannot straightforwardly be related to the difference in narrativity (If the overall frequency development in Face-to-face Conversation was closely connected to narrativity, we would expect a frequency increase mainly driven by increasing use of the past progressive. This, however, is not the case). Eventually, only the past progressive's frequency increase and a relatively small proportion of the present progressive's frequency decrease between ICE-GB and FCSE Face-to-face Conversation can be meaningfully connected to the higher share of narratives in FCSE.

Past progressive As for the past progressive (active) (examples 5.5-7), the frequency increase between ICE-GB and FCSE Face-to-face Conversation has already been discussed. Its overall development in this genre between LLC and FCSE is somewhat less pronounced (and not significant) since a slight decrease occurred between LLC and ICE-GB. It should be noted, though, that the past progressive is generally very frequent in Face-to-face Conversation. Almost certainly, this is a reflection of the fact that Face-to-face Conversation is much more characterised by narrative than the other three genres. Sentence 5.5 serves as an example of progressive use in a narrative passage.

(5.5) *And then off she went, and she* ***was going*** *all over the place.* (FCSE_FFC_17_285)

(5.6) *You* ***were asking*** *the wrong women I think.* (FCSE_BD_6_290)

(5.7) *That's best there, I* ***was just watching*** *the amount of players that* ***were bombing*** *on to support [...].* (FCSE_SC_4_165)

Spontaneous Commentary is the only genre that is characterised by a consistent and significant increase of past progressive use. The development starts from a very low level in LLC (10 tokens), increasing to 29 tokens in ICE-GB and finally reaching the level of 60 tokens in FCSE (since the genres are of similar size, the raw figures are comparable). This

increase seems particularly interesting in the light of Pfaff et al.'s (2013) findings on the increasing use of the so-called *recentness progressive* (cf. chapter 3.3.2 & 3.4.2). There, the past progressive functions as a marker of recent past and is perfective in nature. These uses are characterised by co-occurring adverbials such as *just*, *recently*, or *the other day*. Similarly, Leech (2004: 31f.) mentions uses in which the past progressive refers to fairly recent communicative happenings without expressing a clear aspectual meaning (e.g. *Paula was saying that [...]*; and example 5.6). Since Spontaneous Commentary contains spoken data referring to live events, i.e. deals with what is happening here and now, it only seems logical that past progressive uses are originally not very frequent. Could it be the case, however, that the observed increase of past progressives in ICE-GB and FCSE goes back to the increasing use of the recentness progressive? The pattern would seem well suited to refer to recent utterances and to events that just happened in the course of a match, race, parade, etc. (cf. example 5.7). In order to infer the development, I analysed all past progressive tokens in Spontaneous Commentary. I looked for co-occurring time adverbials and for communication verbs referring to recent, perfective communicative events. Although this procedure comes with a certain degree of subjectivity, it should be faithful enough to arrive at an informed guess as to whether the recentness progressive is a possible driving force behind the increase of the past progressive.

For the LLC data, two of ten past progressives in Spontaneous Commentary clearly express recentness and perfectivity (i.e. 20%). For ICE-GB the rate is 27.6% (eight out of 29 tokens), and for FCSE 30% (18 out of 60 tokens). Given the small sample, this upward – but statistically not significant – trend should not be overestimated. Even if we propose an increasing use of the recentness progressive in Spontaneous Commentary, the increase would not be pronounced enough to account for the observed development of the past progressive in this genre. The recentness progressive is at best one contributing factor among others.

In the remaining genres, the development of the past progressive looks as follows: In Broadcast Interviews & Discussions, a significant increase between LLC & ICE-GB is neutralised by a significant decrease between ICE-GB & FCSE. In Parliamentary Language & Perpared Speech a major decline between LLC & ICE-GB is followed by a slight increase between ICE-GB & FCSE, leading to a clear overall decrease.

Based on these differing, genre-specific developments of the past progressive, it is not possible to identify one underlying trend. There is relatively stable usage in the two conversational genres (Face-to-face Conversation and Broadcast Interviews & Discussions), a clear upward trend in live commentary (which might to a certain degree be based on the increasing use of the recentness progressive), and a decline in the most conservative genres, characterised by scripted speech (Parliamentary Language & Prepared Speech).

Perfect progressive Compared to the present and past progressive, the perfect progressive, i.e. present and past perfect progressive (examples 5.8 & 5.9), is infrequent in spoken English (especially the past perfect is very rare). No clear trend can be made out in the developments of the different genres. A significant and consistent decline can be

observed for Face-to-face Conversation. Usage in Parliamentary Language & Prepared Speech shows a moderate upward tendency. Broadcast Interviews & Discussions are characterised by stable usage between LLC & ICE-GB, compared to a pronounced and significant increase between ICE-GB & FCSE. In Spontaneous Commentary usage is fluctuating, resulting in similar frequency levels for LLC and FCSE.

(5.8) *So over the last few weeks [...] we* ***have been working*** *with the nurse that came back from Sierra Leone.* (FCSE_PS_17_207)

(5.9) *They* ***'d been studying*** *uh Hinduism.* (FCSE_FFC_12_1169)

In the light of these developments one can assume that the perfect progressive is no decisive factor in the progressive's recent frequency increase in spoken British English.

Progressive + modal auxiliaries Matters are different for so-called modal progressives, i.e. progressives that occur in connection with a modal auxiliary (modal + *be -ing*). Broadcast Interviews & Discussions, Parliamentary Language & Prepared Speech, and Spontaneous Commentary, i.e. those genres that exhibit a consistent upward trend of progressive frequencies, are all characterised by increasing use of modal progressives.[12] Usage in Face-to-face Conversation is fairly stable, showing a certain degree of non-significant fluctuation. In general, though, it seems justified to infer increasing usage. This is in line with results of previous studies presented in chapter 3.3.2, reporting a significant frequency increase of modal progressives in British English. It is the combination *will + be -ing*, often used as the so-called matter-of-course future (cf. chapter 3.4.3; example 3.23 & 3.24), that has been identified as the major contributing factor (cf. Leech et al. 2009: 295). In order to check whether this is also the case in the present study, I analysed modal usage in the three subcorpora, conflating the genre distinction (otherwise, the numbers for the different modal verbs would be too small for a meaningful analysis). The results are summarised in Table 5.3.[13]

First, it can be observed that overall use of modal progressives has indeed significantly increased over time. While a non-significant increase of 8.3% has occurred between LLC and ICE-GB, a significant increase of 38.2% has occurred between ICE-GB and FCSE (the increase between LLC and FCSE is significant as well). This corroborates the tentative findings of Smith (2005: 77), who reported increasing use of modal progressives in his two mini corpora of spoken British English.

The increase of 8.3% between LLC and ICE-GB is less pronounced than the increase reported by Leech et al. (2009: 139,295) and Smith (2005: 146) for late 20th-century written English (both 25%). Importantly, though, the change has gained further momentum in spoken language between the 1990s and today, resulting in the pronounced

[12] The trend is significant for Broadcast Interviews & Discussions and Spontaneous Commentary. Cf. Table A.3 for the exact figures and significance levels.

[13] The table comprises nine central modals *can, could, may, might, shall, should, will, would, must* as well as the marginal auxiliaries *need to* and *ought to* and the semi-modal *have to* (cf. Biber et al. 1999: 483f.).

	freq LLC	freq LLC htw	freq ICE	freq ICE htw	freq FCSE	freq FCSE htw	LLC/ICE sign. level	% change LLC/ICE	ICE/FCSE sign. level	% change ICE/FCSE	LLC/FCSE sign. level	% change LLC/FCSE
will	41	11.0	35	10.2	93	20.2	*ns.*	–7.3%	***	+98.0%	***	+83.6%
would	16	4.3	17	4.9	28	6.1	*ns.*	+14.0%	*ns.*	+24.5%	*ns.*	+41.9%
may	11	3.0	6	1.7	5	1.1	*ns.*	–43.3%	*ns.*	–35.3%	*ns.*	–63.3%
should	10	2.7	14	4.1	31	6.7	*ns.*	+51.9%	*ns.*	+63.4%	*	+148.1%
shall	9	2.4	3	0.9	2	0.4	*ns.*	–62.5%	*ns.*	–55.6%	*	–83.3%
must	6	1.6	5	1.5	3	0.7	*ns.*	–6.3%	*ns.*	–53.3%	*ns.*	–56.2%
could	2	0.5	8	2.3	6	1.3	*ns.*	+360.0%	*ns.*	–43.5%	*ns.*	+160.0%
might	2	0.5	6	1.7	10	2.2	*ns.*	+240.0%	*ns.*	+29.4%	*ns.*	+340.0%
have to	1	0.3	2	0.6	2	0.4	*ns.*	+100.0%	*ns.*	–33.3%	*ns.*	+33.3%
ought to	1	0.3	3	0.9	0	0.0	*ns.*	+200.0%	*ns.*	–100.0%	*ns.*	–100.0%
can	0	0.0	0	0.0	1	0.2	*ns.*	NA	*ns.*	Inf	*ns.*	Inf
need to	0	0.0	0	0.0	2	0.4	*ns.*	NA	*ns.*	Inf	*ns.*	Inf
Total	**99**	**26.6**	**99**	**28.8**	**183**	**39.8**	***ns.***	**+8.3%**	**	**+38.2%**	**	**+49.6%**

Table 5.3.: Modal auxiliaries co-occurring with the progressive active (Fisher's exact test: *** p<0.001, ** p<0.01, * p<0.05), htw = hundred thousand words.

increase of 38.2%. This, again, is in line with Smith and Leech (2013: 87), who report an steepening of the frequency increase between F-LOB and BE06.

Overall, modal progressives have increased by almost 50% in spoken British English over the past 60 years. Since they are much less frequent than the present progressive, their impact on the progressive's frequency increase is relatively modest. Nonetheless, they are one contributing factor to the progressive's recent frequency development.

Turning to the individual types, *will* clearly sticks out as the most rapidly developing modal (in absolute terms). While its usage remains almost stable between LLC and ICE-GB (ca. 10 tokens/htw), the frequency doubles between ICE-GB and FCSE (this increase is highly significant). Altogether, the *will be -ing* pattern is 83.6% more frequent today than it was 60 years ago. Apart from this pronounced increase, *will* is by far the most frequently used modal in combination with the progressive. In the light of the findings of Leech et al. (2009) and Smith (2005) for late 20th-century written English, the stable frequencies of *will be -ing* between LLC and ICE-GB are somewhat surprising. Leech et al. report a significant 35% increase of the pattern between LOB and F-LOB; Smith an increase of 33%. Consequently, an increase in spoken language for the same period could have been expected. However, the increase of *will be -ing* in spoken English has only gained momentum in recent decades. It should be noted, though, that the frequencies of *will be -ing* in written English did at no point exceed those in spoken English.

Before analysing the development of *will be -ing* in more detail, two further modal types that show a significant frequency development have to be mentioned. The first is *should*, which – coming from a low level of 2.7 tokens/htw – steadily increases to 4.1 tokens/htw in ICE-GB to 6.7 tokens/htw in FCSE. While none of the individual changes is significant, the overall increase between LLC and FCSE is. Altogether, *should be -ing* increases by 148.1%. This finding is largely in line with the results of Smith (2005: 146) and Leech et al. (2009: 295), who report a doubling of *should be -ing* between LOB and F-LOB (lacking statistical significance, though).

Second, *shall be -ing* shows a significant decrease between LLC and FCSE. Already coming from a low level in LLC (2.4 tokens/htw), its use further declines to 0.9 tokens/htw in ICE-GB and finally falls almost entirely out of use in FCSE (0.4 tokens/htw). Here, we seem to witness a case of statistical obsolescence – most likely tied to an overall decline of *shall* (cf. Leech et al. 2009: 295).

Coming back to *will be -ing*, it has already been explained in chapter 3.3.2 that previous research has connected the pattern's increasing use to the development of the so-called matter-of-course meaning, which is aspectually underspecified and does not imply progressivity (i.e. is not just a combination of the future meaning of *will* and progressivity) (cf. Celle and Smith 2010: 248 and example 3.23 & 3.24). Unfortunately, a functional classification of all 169 *will be -ing* tokens turned out to be unfeasible. Deciding whether a token expresses progressivity or the matter-of-course meaning was in many cases a very ambiguous matter. Examples 5.10 and 5.11 are relatively clear cases: In 5.10 *will be -ing* expresses progressive meaning going on at a point in the future. In 5.11, this future plus progressive meaning is not present. Instead, the future event is presented in its entirety and expected to happen as a matter of course. However, ambiguous cases

such as 5.12 are more common than the straightforward ones. Here it is hard to decide which of the two meanings is prevalent.

(5.10) *Well, we* ***'ll be talking*** *more with Barbara about this in a lot more detail [...] a little later, so stay tuned for that.* (FCSE_BD_7_38)

(5.11) *Next year we* ***'ll be celebrating*** *a hundred years of the House of Windsor.* (FCSE_SC_7_82)

(5.12) *We* ***'ll be introducing*** *the top names in just a moment for the British men as for the British women, a big day.* (FCSE_SC_10_4)

While the matter-of-course reading seems to be the dominant meaning in the data, this result is not reliable enough to derive a sound diachronic trend. Even though it seems likely that the *will be -ing* meaning is undergoing a grammaticalization process towards the aspectually underspecified matter-of-course meaning, the present study cannot exactly pinpoint the state of this development in spoken English.

to-Infinitive + progressive Finally, the least frequent of all morphosyntactic realisations (186 tokens in total) shall be addressed: The progressive as part of a *to*-infinitive construction (*V to be -ing*). The label comprises a number of different uses as exemplified in examples 5.13-16.[14]

(5.13) *[...] the vast majority of the effort* ***appears to be going*** *into constructing methods [...].* (DCPSE:DL-E01)

(5.14) *I pondered on it quite a lot and found it you know realised that I didn't really understand what I was* ***supposed to be doing****.* (DCPSE:DI-B51)

(5.15) *They might but they're certainly* ***going to be getting*** *different mobility opportunities [...].* (FCSE_BI_6_34)

(5.16) *And I would like to thank all the staff there for their great kindness to me if they* ***happen to be listening*** *[...].* (DCPSE:DI-D09)

Similar to the modal progressives, an upward trend can be observed for three of four genres. While usage fluctuates in Face-to-face Conversation (showing a significant increase between LLC and ICE-GB, followed by a smaller but still significant decrease between ICE-GB and FCSE, resulting in a non-significant overall decline between LLC and FCSE), the pattern shows a consistent upward development in all other genres.[15]

[14] The patterns *going to be -ing* and *supposed to be -ing* have been included here. From a structural point, this was an obvious decision. Functionally, however, these uses could be analysed also as (semi-)modal progressives. Since the boundaries are not entirely fixed and previous studies did not discuss these uses as modal progressives, I decided to include them in the diverse group of the *to*-infinitive + progressive pattern.

[15] Cf. Table A.3 for the exact figures.

It has to be noted, though, that these developments take place on a low level. Moreover, the development is only significant in Parliamentary Language & Prepared Speech (overall between LLC and FCSE) where progressives plus *to*-infinitive are non-existent in the LLC data and reach a level of 13 tokens/htw (20 tokens absolute) in FCSE. If the genre distinction is conflated, these individual developments translate into a significant increase between LLC and ICE-GB (from 12.3 tokens/htw to 19.8 tokens/htw), followed by a non-significant decrease to 15.7 tokens/htw in FCSE (the overall increase between LLC and FCSE is not significant).

Based on these numbers, one might assume a positive frequency development of progressives occurring in connection with *to*-infinitives. This would be in line with Smith's (2005: 70) and Leech et al.'s (2009: 288) findings for LOB and F-LOB, as well as with Soete's (2014) observation that use of the *going to be -ing* pattern is increasing (cf. chapter 3.3.2). However, the genre-specific developments should be treated with caution because they are based on low absolute numbers. The positive trend is by no means overwhelmingly strong and does seem to be subject to a certain degree of fluctuation. Eventually, the pattern's recent changes are best considered a side effect of the progressive's overall frequency development rather than an important contributing factor.

In order to back up this tentative conclusion, I decided to make use of the 1.6 billion words Hansard corpus (Alexander and Davies 2015) to chart the diachronic development of progressives + *to*-infinitives in UK Parliamentary Debates. While Hansard clearly has its weaknesses (cf. chapter 4.2.2), it can nonetheless be a valuable supplementary resource, especially with regard to low-frequency items.

I selected the five most frequent *to*-infinitive patterns in DCPSE and FCSE (*going*, *SEEM*[16], *supposed*, *HAPPEN*, *meant*) and searched Hansard for their use between 1900 and the first decade of the 21st century.[17] The results are depicted in Figure 5.4.

SECTION	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
FREQ	360	73	72	995	1204	580	1967	2839	1489	2225	2173
PER MIL	5.57	0.91	1.00	10.45	12.69	4.79	12.94	17.38	8.10	12.56	24.57

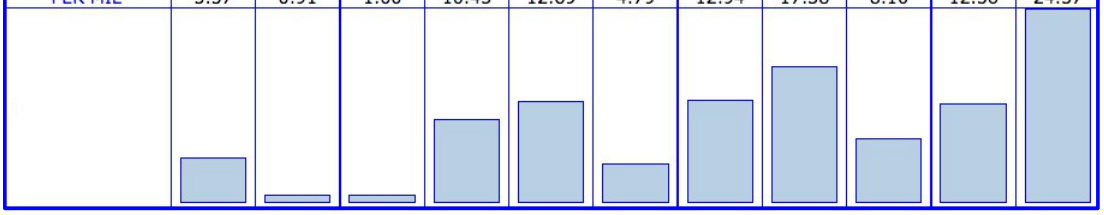

Figure 5.4.: Diachronic development of progressives following *to*-infinitives in the Hansard corpus.

The general tendency seems to confirm the above-made conclusion: Starting from a very low level in the early 20th century, the pattern gets more and more frequent, reaching the highest frequency in the first decade of the 21st century. Similar to my data, the trend is subject to a certain degree of fluctuation, not progressing consistently.

[16] Capitalised words refer to lemmas.

[17] Search term: going|seem|seems|seemed|supposed|happen|happens|happened|meant to be [v?g*].

Importantly, though, the normalised frequencies in Hansard are much lower than the frequencies in ICE-GB and FCSE Parliamentary Language & Prepared Speech. While the *trend* is roughly consistent with my data, the *level of use* remains far below the frequencies presented above. To a certain degree, this is due to the fact that only the five most frequent uses were searched in Hansard. All other uses are not included in the data. Furthermore, the FCSE data is ten years more recent than the Hansard data, which does not go beyond the year 2005. In addition to that, the Hansard data comprises parliamentary debates only, while parliamentary debates and prepared speeches are treated as one genre in DCPSE & FCSE. Considering these points, a certain amount of difference between the usage levels in DCPSE & FCSE and Hansard can be accounted for.

What is more – as explained in chapter 4.2.2 – the Hansard transcripts are not entirely faithful representations of actual speech. They have been edited and standardised with regard to a written norm. Thus, it seems likely that not all instances of the recently emerging progressive + *to*-infinitive pattern have actually been transcribed (for example, *going to be -ing* could have been replaced by the more formal modal construction *will be -ing*; or *supposed to be -ing* by the modal construction *should be -ing*). Consequently, a close comparison between the two results should be avoided. In terms of the general trend, however, the comparison shows that the above-made claims concerning the recent development of progressives in combination with *to*-infinitives seem credible.

Summing up the analysis so far, it can be stated that the present progressive is indeed the major frequency driver behind the progressive's recent increase in spoken British English. It is the most common and prototypical use of the construction and contributes most to its development. As for the past and the perfect progressive, no clear statements can be made. However, there are two uses – modal progressives and progressives occurring in connection with *to*-infinitives – that also contribute to the construction's frequency development. Admittedly, though, their overall contribution per htw is a relatively modest one. Both patterns are no innovations; some of their realisations – such as the pattern *going to be -ing* – are, however, relatively recent additions to the progressive's morphosyntactic paradigm. Modal progressives as well as progressives + *to*-infinitives are clearly less entrenched/prototypical constructional variants than the present progressive. Thus, the results support Hypothesis 2, which assumes a contribution of such uses to the progressive's overall increase, but also expects this contribution to be less pronounced than the increase of the most prototypical of all uses, the present progressive. One aspect of the development can only partially be explained: the decline of the present progressive between ICE-GB and FCSE Face-to-face Conversation. While to a certain degree it can be attributed to an increase in narrative passages and, thus, to increasing use of the past progressive, it is difficult to come up with a conclusive explanation for the remaining part of the development.

Relative frequency development

I would like to conclude the analysis of the progressive's morphosyntactic context by analysing its development from an additional, slightly different angle. After presenting the development per htw (approximating the status in the language system), the analysis now focuses at the relative frequency development of the different morphosyntactic realisations (approximating changes in constructional typicality). How does the development per htw "translate" into the development of the different morphosyntactic realisations relative to each other? Has one morphosyntactic realisation become more or less typical over time; i.e. has the constructional network changed?

As explained in chapter 4.4.2, mosaic plots are a very suitable means to this end. The findings are displayed in Figure 5.5, which comprises four mosaic plots – one for each genre. While the size of the individual tiles corresponds to the token frequency of the respective morphosyntactic realisation, the colouring indicates whether a variable is over- (blue) or underrepresented (red).

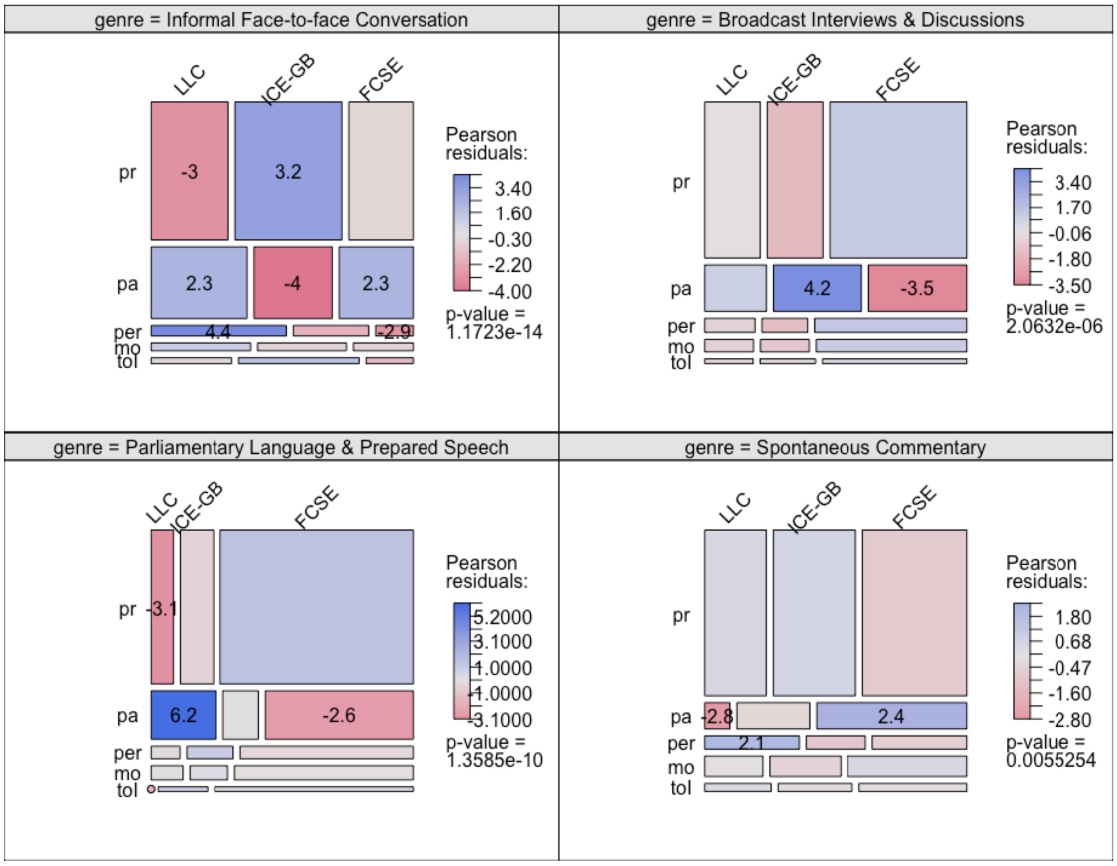

Figure 5.5.: Morphosyntactic context by corpus and genre (active progressives only). pr = present, pa = past, per = present & past perfect, mo = modal, toI = *to*-Infinitive. Numbers and shading indicate Pearson Residuals.

As for the present progressive, the previous section revealed that it has been the major frequency driver in Broadcast Interviews & Discussions, Parliamentary Language & Prepared Speech, and Spontaneous Commentary. In Face-to-face Conversation, the present progressive is more frequent in FCSE than in LLC, its peak, however, occurs in ICE-GB. In the respective mosaic plots, this plays out as follows: In Face-to-face Conversation, the present progressive is significantly overrepresented in ICE-GB (+3.2), while it is underrepresented in LLC (–3)[18]. In FCSE, it is neither especially typical nor atypical compared to the other morphosyntactic variants. This means that the most prototypical core use of the progressive construction was more typical of progressive use in Face-to-face Conversation in the 1990s than it is today.

In Broadcast Interviews & Discussions, no significant over- or underrepresentation of the present progressive can be observed. Importantly, though, some non-significant developments have occurred: The present progressive is more frequent than expected in FCSE, while it is less frequent than expected in ICE-GB. Basically no deviation can be reported for the LLC data. The slight underrepresentation in ICE-GB is a reflection of the stable frequencies per htw between LLC and ICE-GB and the simultaneous increase of other uses (esp. past progressive). More interestingly, though, the pronounced frequency increase of the present progressive per htw between ICE-GB and FCSE has not resulted in a significant increase in typicality, which is due to the fact that other morphosyntactic realisations (perfect, modals, *to*-infinitives) have taken a positive development as well. These results show how the different frequency metrics shed light on different aspects of the development that would otherwise go unnoticed.

In Parliamentary Language & Prepared Speech, the plot reveals a significant underrepresentation of the present progressive in LLC (–3.1). In ICE-GB it has reached a level very close to the expected frequency. In FCSE it is more frequent than expected; however, without reaching a significant residual level above 2. Altogether, though, the present progressive's increase per htw is mirrored by a steady increase in typicality compared to the other morphosyntactic realisations.

Finally, in Spontaneous Commentary we get very light colouring (i.e. small residuals) for each of the three tiles or the present progressive. This means that no major relative frequency shifts have occurred. It is interesting, however, that the present progressive is slightly overrepresented in LLC and ICE-GB while being slightly underrepresented in FCSE. At first glance this seems to be at odds with the observed development per htw. However, the phenomenon is explained by the fact that alongside the present progressive, other uses (most importantly the past progressive) have increased in an even more pronounced way (not in terms of the absolute contribution to the overall increase but in terms of percental growth (cf. Table A.3)). Thus, the pronounced increase (per htw) of the present progressive in Spontaneous Commentary has not resulted in an increase in typicality compared to the other progressive uses. Put differently, the present progressive has become more entrenched in the language system over time, but it has not become more typical of progressive use.

[18] As explained in chapter 4.4.2, residuals above/below +/–2 indicate a statistically significant deviation from the expected frequencies.

Considerable variation can be observed for the past progressive. In Face-to-face Conversation, its use is significantly overrepresented in in LLC (+2.3) and FCSE (+2.3), while being significantly underrepresented in ICE-GB (–4). Thus, no clear diachronic trend can be made out. The clear underrepresentation of ICE-GB past progressives is due to the fact that the pattern declines between LLC and ICE-GB (per htw), while the present progressive massively increases at the same time (per htw). A slight increase between ICE-GB and FCSE (per htw), linked to a decrease of the present progressive (per htw) are the major reasons for the positive residual in FCSE Face-to-face Conversation. The plot suggests that the relative frequency developments of the present and the past progressive work in opposite directions: As the present progressive becomes more typical, the past progressive gets less typical (and vice versa) (this – in general – is also the case in the remaining three genres).

In Broadcast Interviews & Discussions, the past progressive is significantly overrepresented in ICE-GB (+4.2), which is mirrored by a (non-significant) underrepresentation of the present progressive. In FCSE, the past progressive is significantly less frequent than expected (–3.5), while the present progressive is non-significantly overrepresented. Only a very small degree of variation can be observed for the LLC data. The plot does not reveal a consistent trend regarding the past progressive's relative frequency in Broadcast Interviews & Discussions.

This is different for Parliamentary Language & Prepared Speech. Here, the past progressive exhibits a consistent negative trend: While it is very strongly and significantly overrepresented in the LLC data (+6.2), use in ICE-GB is as frequent as expected. Finally, FCSE past progressives are significantly underrepresented (–2.6). The present progressive shows exactly the reverse trend, from under- to overrepresentation.

Opposite to the development in Parliamentary Language & Prepared Speech, past progressive use in Spontaneous Commentary develops from significant underrepresentation in LLC (–2.8) to significant overrepresentation in FCSE (+2.4). This relative frequency shift goes back to the pronounced frequency increase of the past progressive per htw. Even though the present progressive has also significantly increased in frequency per htw, the increase of the past progressive is still more pronounced. Thus, it is the past progressive that has become more typical over time and not the present progressive.

Summing up the relative frequency development of the past progressive, two observations have to be highlighted: First, usage of the past progressive is subject to a considerable degree of variation, which is reflected in comparatively high/low residuals. Second, this variation does not seem to follow a consistent trend. While the past progressive has become more typical of progressive use in Spontaneous Commentary, the opposite is the case for Parliamentary Language & Prepared Speech and Broadcast Interviews & Discussions. Still different is the development in Face-to-face Conversations, where past progressive usage goes from over- to under- and back to over-representation. Put differently, the high degree of variation indicates shifting constructional preferences, while the absence of a clear trend makes it difficult to come up with a generally valid interpretation of these genre-specific changes. Perhaps the clearest underlying phenomenon is the opposite development of present and past progressive: As the past progressive becomes more typical, the present progressive gets less typical (and vice versa).

Of the remaining three morphosyntactic realisations, only the perfect progressive exhibits some significant variation. In Face-to-face Conversation, LLC perfect progressives are significantly overrepresented (+4.4), while FCSE perfect progressives are significantly underrepresented (–2.9); the ICE-GB tokens occupy a middle position. Thus, the decline per htw also translates into reduced constructional typicality.

Very little variation can be reported for Broadcast Interviews & Discussions and Parliamentary Language & Prepared Speech. In Spontaneous Commentary, however, LLC perfect progressives are significantly more frequent than expected and stick out as typical (+2.1). In ICE-GB and FCSE, this status is lost and perfect progressives become very slightly underrepresented, thus showing a similar development as in Face-to-face Conversation.

Modal progressives and *to*-infinitives only show minor variation, with none of the residuals reaching a level above/below two. This means that their relative share among the other morphosyntactic realisations has not changed significantly. In the light of their positive development per htw (esp. for modal progressives) this is an interesting insight: While modal progressives have become significantly more frequent in three of four genres, these developments are not strong enough to make them significantly more typical of progressive use. Exactly the same is true for progressives occurring with a *to*-infinitive construction.

In summary, the analysis of the relative frequency development reveals that increased corpus frequencies per htw do not consistently result in increased constructional typicality of the respective morphosyntactic realisations. This can be the case (as for the present progressive in Parliamentary Language & Prepared Speech), but does not necessarily have to be so (as for the present progressive in Spontaneous Commentary) and crucially depends on the development of the whole morphosyntactic paradigm.

While general trends could be inferred from the development per htw, the situation is less clear with regard to relative frequencies and the progressive's constructional network.

5.2.2. Verbal development

The second central aspect of progressive use apart from the different morphosyntactic realisations is the construction's verbal and semantic paradigm. As has been reported in chapter 3.3.2, previous corpus findings suggest a modest increase of progressive use with stative verbs in the course of the 20th century. In this context, it has been proposed that the progressive might be developing into a general imperfective marker. This and other claims will be addressed in this subchapter.

First, the frequency development of different verb classes will be analysed (in the same way as the different morphosyntactic realisations have been analysed). Second, Multiple Distinctive Collexeme Analysis will be used to analyse the diachronic development of individual progressive verbs.

Verb classes: Development per hundred thousand words

In order to approximate general tendencies of the progressive's semantic development, I decided to classify all verbs according to eight semantic classes: activity/event (e.g. *go, do, happen*), process/aspectual (e.g. *develop, continue, change*), communication (e.g. *talk, say, joke*), cognition/emotion/attitude (e.g. *think, love, hope*), perception/sensation (e.g. *look, feel, see*), causative (e.g. *allow, enable, force*), stance (e.g. *sit, live, wait*), and existence/relationship (e.g. *be, have, seem, contain*). This classification is based on the verb classes proposed in Biber et al. (1999: 360ff.) and Leech (2004: 23ff.). While it is clear that such a classification is to some extent subjective (e.g. What are suitable verb classes? Which verb belongs to which class?) and has to accept a certain degree of flexibility (esp. with regard to polysemous verbs defying unambiguous classification[19]), it comes with the major advantage of bringing to light the general semantic preferences and developments of the progressive.

Regarding a dynamic/stative distinction, the eight semantic classes can be further grouped as follows: activity/event, process/aspectual and communication verbs clearly refer to dynamic situations and, consequently, can be summarised as dynamic verbs – freely combining with the progressive. Existence/relationship and stance verbs refer to stative siuations and can be called state verbs. Finally, cognition/emotion/attitude, perception/sensation and causative verbs somehow fall in between the dynamic/stative categories. Their use with the progressive includes ambiguous cases such as *Our politics are enabling a smooth procedure*, where it is hard to decide whether the causative verb *enable* refers to an active or stative situation. Another example is *think* as in *I'm thinking about emigrating to Canada.* The situation can be classified as a temporary mental state but also as a mental activity. With regard to perception/sensation verbs, *look* is an especially tricky case. While it can be clearly active (*He's looking towards you*), it can also refer to stative situations (*You're looking good!*). Finally, there are examples that defy clear classification and have both qualities of a state and an activity (e.g. *He has been looking in this direction for hours*). On the basis of these considerations, cognition/emotion/attitude, perception/sensation and causative verbs are best regarded as intermediate verbs on an dynamic/stative scale.

Figure 5.6 visualises the normalised progressive frequencies of each verb class for each of the four genres over time[20] (the numerical results are summarised in Table A.4; Figure A.7 shows the development per htw according to the active/dynamic/stative distinction).

Activity/event The clearest aspect of the plot is the high frequencies of activity/event verbs (cf. examples 5.17-19). Much in the same way as the present progressive is the dominant morphosyntactic realisation, activity/event verbs form the semantic backbone

[19] *have*, for example, cannot only be used statively, but also with dynamic meaning (*We're having a party*). Nonetheless, it was only assigned to one semantic class – existence/relationship. While such a classification procedure results in a certain degree of noise in the data, it was nonetheless preferred over classifying each and every single token manually, which would have been extremely time-consuming.

[20] A display without activity/event verbs is provided in Figure A.6 in the appendix. There, changes in the less frequent verb classes become more obvious.

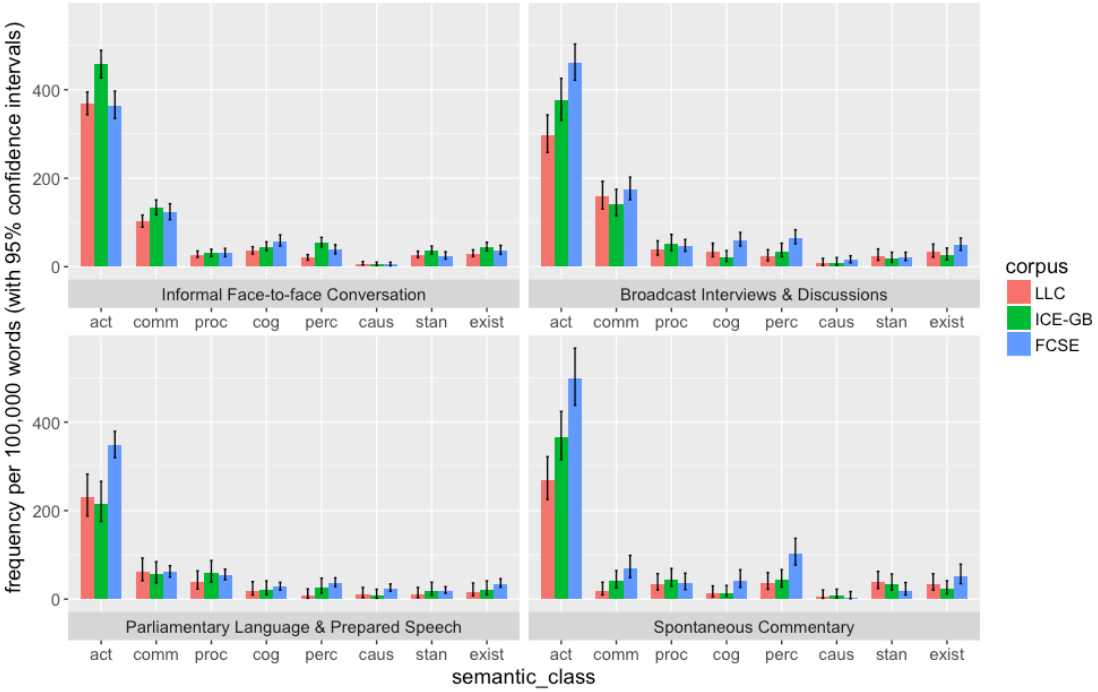

Figure 5.6.: Verb class by corpus and genre (active progressives only). act = activity/event, caus = causative, cog = cognition/emotion/attitude, comm = communication, exist = existence/relationship, perc = perception/sensation, proc = process/aspectual, stan = stance.

of progressive use. Considering previous research and the fact that activity/event verbs are the largest verb class in English, this does not come as a surprise.[21] It corroborates the claim that activity/event verbs are much more prototypical of progressive use – and much more routinised – than for example the small class of causative verbs.

(5.17) *She**'s probably doing** more than that I think.* (FCSE_BD_6_149)

(5.18) *And I **was working** in an office in Liverpool [...].* (FCSE_FFC_12_268)

(5.19) *[...] inequality **is actually hitting** our economy [...].* (FCSE_PL_10_146)

Let us now focus on the diachronic development. In Face-to-face Conversation, frequencies for activity/event verbs between LLC and FCSE are stable. However, this does not mean that no change has occurred. Similar to the development of the present

[21] Studying progressive use in different world Englishes, Rautionaho (2014: 214) shows that activity and communication verbs account for the vast majority of all progressive uses. Similarly, also studying progressive use in different global varieties of English, Deshors (2017: 278) observes that stative uses of the progressive are rare, which is in line with the results of Leech et al. (2009: 292) presented in chapter 3.3.2.

progressive, use of activity/event verbs significantly increases between LLC and ICE-GB, followed by a significant decline between ICE-GB and FCSE, almost mirroring the overall frequency development of the progressive in this genre.

In the remaining genres, a different picture emerges: In Broadcast Interviews & Discussions, activity/event verb progressives significantly increase by 26% between LLC and ICE-GB, followed by a significant increase of 22.7% between ICE-GB and FCSE. Even more pronounced is the increase in Spontaneous Commentary (LLC - ICE-GB: +35.8%, ICE-GB - FCSE: +36.4%). In Parliamentary Language & Prepared Speech, a non-significant decline of 6% between LLC and ICE-GB is followed by a major increase of 60.9% between ICE-GB and FCSE. In all of these genres, overall progressive frequencies have also increased significantly.

Thus, like the present progressive, progressives occurring with activity/event verbs are a major force behind the progressive's frequency development. Quantifying the growth between LLC and FCSE in tokens per htw, a minor decline of 4 tokens/htw in Face-to-face Conversation stands against major gains of 163 tokens/htw in Broadcast Interviews & Discussions, 118 tokens/htw in Parliamentary Language & Prepared Speech, and 229 tokens/htw in Spontaneous Commentary.

Process/aspectual The second verb class qualifying as dynamic is the much smaller class of process/aspectual verbs (cf. examples 5.20-21). As the name already indicates, these verbs readily combine with the progressive construction and, from a semantic point of view, there is no reason to regard them as less prototypical progressive verbs than activity/event verbs. However, from a corpus perspective (less frequent), they are less prominent examples of progressive usage.

(5.20) *And yet, something very un-British* ***is taking root*** *in our politics.* (FCSE_PS_1_27)

(5.21) *[...] the nature of the steel industry* ***is changing*** *[...].* (FCSE_PL_5_141)

Diachronically, they exhibit an interesting development: Between LLC and ICE-GB, their use with the progressive increases in all genres. The rates lie between 9.8% and 51.6%. However, the development does not reach statistical significance in any of the genres. This positive development comes to a halt or is reversed in all genres in the next diachronic stage from ICE-GB to FCSE (rates between +1.7% and –20.3%). Again, none of the developments reaches statistical significance. Overall, from LLC to FCSE, all genres exhibit a growth between 3.8% and 41.4% (below significance). On the basis of these findings – which lack statistical significance but are similar in their developmental course – one might assume a generally positive frequency trend with a usage plateau in the 1990s.[22]

[22] It is important, though, to treat such tentative trends with some caution. First, it is always possible that they are due to chance. Second, they might simply be reflections of larger trends. For example, it is possible that the overall frequency of process/aspectual verbs has changed over time and that the trend witnessed for their combination with the progressive is simply a reflection of this development. In this case, the increased progressive use would still be a phenomenon in its own right, but the interpretation of the results would be a different one.

Communication The last verb class qualifying as dynamic is communication verbs, which, strictly speaking, can be regarded as a subcategory of activity/event verbs (communicative activities). Sentences 5.22-24 exemplify their use in connection with the progressive.

(5.22) *I mean **are you talking** about halving it?* (FCSE_BI_11_74)

(5.23) *So, hold on a second, so you**'re saying** if the public are misinformed and the politicians simply reflect that, that's okay?* (FCSE_BI_12_30)

(5.24) *Sorry Tez, I **was just saying** that uh (,) it's a good job that Leeds found themselves with uh (,) the centre, Ben Currie.* (FCSE_SC_6_112)

Progressive use with communication verbs shows the clearest development in Spontaneous Commentary. The frequency increases by 110.8% (+22 tokens/htw) from LLC to ICE-GB. With a plus of 69.7% (+29 tokens/htw) also the increase from ICE-GB to FCSE is very pronounced. The overall increase between LLC and FCSE (+257.7%) is highly significant.

If one also considers the fact that past progressive use has dramatically increased in Spontaneous Commentary and that use of the so-called recentness progressive is frequently connected with communication verbs (i.e. recent communicative activities), it is tempting to read this development as a further hint towards increasing usage of the recentness progressive in the genre of Spontaneous Commentary (cf. the discussion in chapter 5.2.1). This assumption is backed by cross-tabulating and normalising the frequencies of past & communication verb progressives: their frequency increases from 8.6 tokens/htw in LLC to 14.3 tokens/htw in ICE-GB, reaching 29.5 tokens in FCSE. Admittedly, though, the raw figures of this pattern in Spontaneous Commentary are very low (LLC = 4, ICE-GB = 7, FCSE = 14), not allowing far-reaching conclusions. Nonetheless, they are generally in line with the above-mentioned idea that increasing use of the recentness progressive might be a phenomenon characteristic of Spontaneous Commentary. The progressive in example 5.24 – uttered by the commentator of a rugby match – serves as a nice example of how past progressives referring back to recent communicative activities can be used in this genre.

In Face-to-face Conversation and Broadcast Interviews & Discussions, progressives occurring with communication verbs are frequently used (esp. the frequencies in Broadcast Interviews & Discussions are very high) and also show a positive development between LLC and FCSE. However, the increase is comparatively small and the trends are not linear. The overall increase between LLC and FCSE is much less pronounced than in Spontaneous Commentary and does not reach statistical significance in either of the genres (Face-to-face Conversation: +20.4%, Broadcast Interviews & Discussions: +10.3%).

The especially frequent use of communication verbs in connection with the progressive in Broadcast Interviews & Discussions (between 142 and 175 tokens/htw) is reflected in examples 5.22 and 5.23. Both sentences stem from interview situations, and in both examples the speaker refers back to an utterance of his/her interlocutor. Both progressives

co-occur with the personal pronoun *you* and are part of a question that can be understood as an interpretation of the interlocutor's previous utterance. Such interpretative progressives are subjective in nature, which is underlined by the use of the discourse marker *I mean* and the sequence *hold on a second*. In both cases, progressive use can be seen as a communicative strategy by which a previous utterance of the interlocutor is taken up, interpreted and used for a subsequent question. In modified form, this pattern is frequently found in Broadcast Interviews & Discussions and – at least partially – helps to explain the frequent use of communication verbs in connection with the progressive. The observation is in line with Smith's (2005: 92) insight that interpretative progressives frequently co-occur with communication verbs (cf. chapter 3.4.2; example 3.15).

The only genre that shows no increasing use of progressives plus communication verbs is Parliamentary Language & Prepared Speech. Numbers in all three subcorpora are fairly similar and the frequency in LLC (62 tokens/htw) is almost equal to the one observed in FCSE (61 tokens/htw).

Figure A.7 (appendix) visualises the joint development (per htw) of the three dynamic verb classes (activity/event, process/aspectual, communication). While usage fluctuates in Face-to-face Conversation, very pronounced upward trends have occurred in the remaining three genres.

Cognition/emotion/attitude An even clearer upward development than for communication verbs is observed for cognition/emotion/attitude verbs (cf. examples 5.25-28).[23] Progressives occurring in combination with this verb class are considerably less frequent than progressives with communication verbs. Highest frequencies are found for Face-to-face Conversation and Broadcast Interviews & Discussions, while usage is on a lower level in Spontaneous Commentary and Parliamentary Language & Prepared Speech. Thus, progressives co-occurring with cognition verbs seem to be a feature of conversational style. Given the fact that the present work regards this verb class as intermediate on a dynamic/stative scale, i.e. as non-prototypical progressive verbs, a preference for informal/conversational style does not come as a surprise.

(5.25) *And it's come to the point now where we* ***'re thinking*** *about getting an electric wheelchair for him.* (FCSE_FFC_8_879)

(5.26) *And people* ***are thinking****, should that not be the other way around?* (FCSE_FFC_10_870)

(5.27) *[...]* ***I'm just admiring*** *your jewellery.* (FCSE_FFC_12_480)

(5.28) ***I'm liking*** *your clothes, this is not good for the podcast but you're very Christmassy.* (FCSE_BD_8_61)

A comparison of their normalised frequencies in LLC and FCSE reveals clear upward trends in all four genres. In three genres, this development is statistically significant

[23] From now on, I will refer to this class simply as *cognition* verbs.

(Face-to-face Conversation: +60.5%, +22 tokens/htw; Broadcast Interviews & Discussions: +73%, +25 tokens/htw; Spontaneous Commentary: +226.4%, +29 tokens/htw). Only the development in Parliamentary Language & Prepared Speech does not reach a significant level (+44.8%, +9 tokens/htw). Regarding the course of the development, three of four genres (Face-to-face Conversation, Parliamentary Language & Prepared Speech, Spontaneous Commentary) exhibit a steady upward development. In Broadcast Interviews & Discussions, a non-significant decline between LLC and ICE-GB is followed by a highly significant increase between ICE-GB and FCSE.

Based on these figures, we can conclude that use of the progressive with cognition verbs is more routinised today than it was in the 1950s.

If we consider the progressives in sentences 5.25 and 5.26, we see that the one in 5.25 refers to a mental process/activity – the speaker refers to a decision process that has not come to an end yet. In 5.26, however, the same verb, *think*, does not refer to a clearly dynamic situation. While the act of thinking might theoretically also involve a reflection process in this example, the situation actually seems to refer to a state of mind/an opinion that people hold. This is reflected in the fact that use of the simple aspect seems perfectly acceptable – if not more natural – in this example. Similarly, the progressive in example 5.27 also refers to a temporary state of mind rather than an active act of admiration. It is only the act of explicitly verbalising it that lends the situation a dynamic touch. The progressive in 5.28 is the least prototypical and most salient example. *I like your clothes* would clearly sound more natural. One might even argue that the example constitutes a performance error, since – from a grammatical point of view – there is no reason to use the progressive aspect. However, it is likely that its use is pragmatically licensed and serves an emphasis/foregounding function (cf. chapter 3.4.2).

In the light of these examples, the classification of cognition verbs as intermediate on the dynamic/stative continuum seems justified. While the class includes tokens that would also qualify as either dynamic (5.25) or stative (5.28), most cases (5.26-27) are less clear-cut and exhibit qualities of both dynamic and stative situations. Their use with the progressive is acceptable but not obligatory. The fact that this verb class' use with the progressive is increasing in all genres can be read as a clear sign that the progressive's semantic scope is widening in the course of its frequency increase. Tellingly, usage is most advanced in the two conversational genres (Face-to-face Conversation and Broadcast Interviews & Discussions). The only genre in which the increase does not reach statistical significance is the most conservative of all: Parliamentary Language & Prepared Speech. This is in line with the assumptions of colloquialization, proposing a spread of informal, non-standard forms from more to less formal genres.

It has to be noted, though, that the overall contribution of this verb class to the progressive's frequency increase is a relatively modest one. In those genres in which the progressive's overall frequency increases, it is a noticeable but no decisive force. Interestingly, also in Face-to-face Conversation, where progressive frequencies today are on the same level as 60 years ago, progressive use with cognition verbs has become significantly more frequent.

Perception/sensation The second class of verbs qualifying as intermediate on the dynamic/stative continuum is perception/sensation verbs. These verbs make reference to perceptual/sensational states and activities. Their use with the progressive is highlighted in the following examples:

(5.29) *He**'s looking** for a free kick there, and he's got one.* (FCSE_SC_2_103)

(5.30) *It's got new doors, it**'s looking** wonderful.* (FCSE_PS_19_185)

(5.31) *Well, you wouldn't, you wouldn't believe it's true, it's hard to believe what we**'re seeing**.* (FCSE_SC_2_539)

(5.32) *But then it would, kind of felt like, why **am I feeling** like that?* (FCSE_FFC_11_382)

While the progressive in example 5.29 refers to a perceptual activity, the one in example 5.30 – also formed with the verb *look* – clearly refers to a state. Thus, perception/sensation verbs include polysemous verbs such as *look* that can be used both with dynamic and stative meaning. The progressive in sentence 5.31 is also stative in nature. It uses the verb *see* to refer to a temporary perceptual state that does not comprise a dynamic element. Consequently, the verb phrase could also be rendered in the simple aspect. Use of the progressive seems to be licensed by two semantic characteristics of the situation: temporariness and incompletion. The third major semantic aspect of aspectual progressive use, dynamicity, is not present. Thus, in such cases, the progressive functions as a marker of imperfective but not of progressive aspect. The same interpretation applies to the progressive in example 5.32, in which *feel* refers to a temporary state. Examples 5.31 and 5.32 are contexts where progressive use is possible but not obligatory. More noteworthy than these uses are examples like 5.30, where even a clear temporary element is missing. From a semantic point of view, the default option in such situations clearly is the simple aspect. Again, however, if pragmatics are taken into account, the choice of the progressive can be explained: it serves as a foregrounding device, lending emphasis to the statement. Thus, while the progressive in example 5.29 has aspectual meaning, the one in 5.30 is pragmatically motivated, and the ones in 5.31 and 5.32 are also aspectual but only express imperfective aspect, lacking progressive meaning.

These considerations show that perception/sensation verbs entail heterogeneous use of the progressive and cannot simply be classified as dynamic or stative. It is justified to call them intermediate not only with respect to the dynamic/stative scale, but also with regard to the question whether they form a prototypical or non-prototypical class of progressive verbs.

The course of their recent diachronic development is unambiguous: Comparing the frequencies of LLC and FCSE, all genres exhibit a statistically significant increase. Except for Face-to-face Conversation – where a sharp increase between LLC and ICE-GB (+169.3%, +34 tokens/htw) is followed by a smaller decline of 30% (–16 tokens/htw) between ICE-GB and FCSE – the increase follows a steady upward trend. In Parliamentary Language & Prepared Speech, usage comes from a very low level in LLC (7 tokens/htw),

reaching a much higher level in ICE-GB (26 tokens/htw) and FCSE (37 tokens/htw). Thus, progressive use with perception/sensation verbs has permeated into the most conservative of spoken genres in the course of the past six decades. In Spontaneous Commentary, a modest increase between LLC and ICE-GB (+17.2%, +6 tokens/htw) is followed by a dramatic one between ICE-GB and FCSE (+140.3%, +60 tokens/htw). In Broadcast Interviews & Discussions, an increase of 53.3% (+12 tokens/htw) is followed by one of 89.1% (+31 tokens/htw).

Together with the increasing use of cognition verbs, this increase of perception/sensation verbs is a clear indication that the progressive's semantic boundaries have widened over the past decades. Since progressive use with these two verb classes comes from a fairly low level, the observed increase does not mean that these less prototypical uses reach levels somewhere near the most dominant of all verb classes, activity/event verbs. However, the developments reflect an increasing readiness to use the progressive in contexts in which its use is not obligatory but optional. As a result, general imperfective uses of the construction might be gaining ground. Importantly, though, this development is no replacement of the prototypical, dynamic uses of the construction: Progressives with activity/event verbs are on the increase in three of four genres. The same is true for communication verbs. Consequently, as the functional core uses of the construction become further entrenched in language use, less prototypical uses follow in the wake of this development, leading to a slow but steady expansion of the verbal paradigm.

Causative The last verb class that qualifies as intermediate on the dynamic/stative scale is causative verbs. It is a small class, which only accounts for a minor fraction of progressive uses. Many of the tokens do not straightforwardly refer to activities or states, but to situations that have qualities of both. This observation is reflected in the following examples:

(5.33) *Why **is he allowing** it to be balkanised?* (FCSE_PL_1_165)

(5.34) *The work of these scientists **is helping** to support initiatives to expand the donor pool [...].* (FCSE_PS_13_219)

(5.35) *[...] it's the Tories' plan to slash tax credits that **is causing** alarm in households across the country.* (FCSE_PS_5_121)

None of the situations referred to is clearly dynamic. In example 5.33, uttered in a House of Commons debate, the person supposedly allowing balkanisation is most likely not pursuing this aim actively. The question is an attack on the political opponent and makes reference to a state of affairs that is undesirable in the eyes of the speaker. Similarly, the progressive in example 5.34, which is part of a lecture, also makes reference to a state of affairs. It has an active element – the scientific work is contributing to a real-world change – but does not entail any additional activities on behalf of the scientists. The same interpretation applies to the last example (5.35): A political plan of the Conservative party results in a state of alarm among certain parts of society. Although the plan comes with real-world consequences, it does not actively seek to bring them

about. As a result of these ambiguities, the verb phrase in all three example sentences could also be coded in the simple aspect. Again, progressive use is possible but not obligatory.

Diachronically, no clear trend emerges. In Face-to-face Conversation, progressive use with causative verbs is extremely infrequent and slowly declines over time. However, this decline is not significant. Usage in Broadcast Interviews & Discussions almost doubles but is still on a low level in FCSE. The development does not reach significance. Spontaneous Commentary is characterised by fluctuating use: a non-significant increase is followed by a non-significant decrease. A significant trend only occurs in Parliamentary Language & Prepared Speech, where a minor decline between LLC and ICE-GB is followed by a more than twofold increase (+18 tokens/htw) between ICE-GB and FCSE. While the raw frequencies are very low (LLC = 4, ICE-GB = 3, FCSE = 38), the numbers might nonetheless be taken as a further hint that readiness to use the progressive in non-obligatory contexts is growing in Parliamentary Language & Prepared Speech.

Taken together, cognition, perception/sensation and causative verbs – i.e. the three intermediate verb classes – show a clear upward trend in all four genres. This development is summarised in Figure A.7.

Stance The remaining two verb classes – stance and existence/relationship – can clearly be considered stative. Stance verbs make reference to static (often physical) situations. Their use with the progressive is demonstrated in the following examples:

(5.36) *Who **is waiting** there to receive that baton?* (FCSE_PS_12_26)

(5.37) *She came home last night, Ferg and I **were sitting** in the living room with Hannah.* (FCSE_FFC_3_461)

(5.38) *We**'re living** together happily and we have three cats and it's all fine.* (FCSE_BD_5_4)

The progressives in the first two examples refer to temporary states. In the second example, the progressive also serves a temporal framing function: The moment she came home last night, Ferg and the speaker were sitting in the living room. Thus, the state of sitting frames the event of coming home.

In 5.38, progressive use cannot straightforwardly be explained with temporariness. Neither does temporal framing serve as an explanation. Instead, the described situation seems to lack any clear aspectual characteristics justifying progressive encoding. Nonetheless, the example seems perfectly acceptable. This is due to the fact that *live* and several other stative verbs (*lurk, wait, stand, wear, hold, stay*) do actually combine with the progressive relatively freely (cf. Biber et al. 1999: 472), counter to the commonly heard claim that stative verbs were *anti*-progressive verbs. Consequently, uses such as 5.38 do come across as natural. Use of the respective verbs with the progressive

is sufficiently routinised to license examples lacking clear semantic motivation. Put differently: Frequent co-occurrence leads to increased acceptability and fewer contextual constraints.

Diachronically, progressive use with stance verbs does show no clear trend. In none of the four genres can a significant development be observed between LLC and FCSE. In Face-to-face Conversation, a non-significant increase of 35.9% between LLC and ICE-GB is followed by a significant decline of 34.6% between ICE-GB and FCSE. Usage in Broadcast Interviews & Discussions also fluctuates, showing a 24.8% decline between LLC and ICE-GB, followed by a 17% increase between ICE-GB and FCSE. Neither of the two developments is statistically significant. In Parliamentary Language & Prepared Speech, a major but still non-significant increase of 93.8% from LLC to ICE-GB results in relatively stable usage between ICE-GB and FCSE (+4.3%). A continuous but also non-significant decline has occurred in Spontaneous Commentary (LLC - ICE-GB: –10.6%, ICE-GB - FCSE: –45.5%).

Altogether, no clear development emerges. Consequently, it seems sensible to interpret the changing frequencies as mere fluctuation in the data and to assume no major diachronic trend for progressive use with stance verbs.

Existence/relationship The second class of verbs qualifying as stative is existence/ relationship verbs. As Biber et al. (1999: 364) point out, "[v]erbs of existence or relationship report a state that exists between entities". Some of the most common representatives of this class are copular verbs (e.g. *be, seem*), others report a particular state of existence (e.g. *exist*), or a particular relationship between entities (e.g. *represent*) (ibid.). In many cases, these verbs do not readily combine with the progressive. This, however, does not mean that no progressive uses can be found in the data. On the contrary, since the class includes the very frequent verbs *be* and *have*, the normalised progressive frequency of existence/relationship verbs is not as low as might be expected. Depending on subcorpus and genre, it lies between 17 and 53 token/htw. Compared to – for example – causative verbs (4 to 25 tokens/htw), this is comparatively frequent. Let us consider the following examples:

(5.39) *My dad being (,) a man's man, he was saying, no son of mine* ***is being*** *a hairdresser [...].* (FCSE_FFC_19_96)

(5.40) *And they* ***weren't being*** *vindictive, I think they were trying to be protective, but uh.* (FCSE_FFC_15_101)

(5.41) *The police came down and said [...] you uh, you****'re having*** *an illegal rave [...].* (FCSE_FFC_17_18)

(5.42) *So I****'d been having*** *a few problems [...].* (FCSE_BD_6_104)

(5.43) *But actually, you don't get your space with a child like Hannah, 'cause you****'re all constantly having*** *to make those decisions [...].* (FCSE_FFC_5_890)

(5.44) *Oh, she****'s wearing*** *a scarf.* (FCSE_FFC_13_188)

The progressive with *be* in example 5.39 is a futurate progressive (cf. chapter 3.4.3). It refers to the son's plan of being a hairdresser, which is rejected by the father.

The one in example 5.40 serves a different function: it follows the pattern *BE being adj.* (cf. chapter 3.3.2), which often indicates the attribution of deliberate behaviour. While *they weren't vindictive* would refer to stable character traits, *they weren't being vindictive* offers an evaluation/interpretation of a specific, temporary behaviour (cf. Ljung 1980: 69ff.). In this sense, progressive *be* is not used statively but dynamically. It has been suggested that this pattern of progressive use has recently increased in frequency. I will come back to this question below.

The progressives in sentences 5.41-43 are all formed with *have*. However, in each example a different meaning can be discerned. In 5.41, *have* is used dynamically, referring to the act of conducting a rave. Thus, as in example 5.40, not all progressives that are formed with an existence/relationship verb do actually qualify as stative. Counter to this, progressive *have* in example 5.42 refers to a temporary, stative situation. Still different is the use in example 5.43, where progressive *have* is used as a semi-modal with deontic meaning.

Progressive use of other existence/relationship verbs is exemplified in example 5.44, where progressive *wear* refers to a temporary state.

Diachronically, a comparison of the genre-specific LLC and FCSE frequencies suggests increasing progressive use of existence/relationship verbs. However, unlike for cognition or perception/sensation verbs, this trend is less solid and does not reach statistical significance in any of the genres. While frequencies in all genres are higher in FCSE than in LLC, a continuous upward trend from LLC to ICE-GB to FCSE can only be observed for Parliamentary Language & Prepared Speech. In Face-to-face Conversation, a significant upward development between LLC and ICE-GB (+48.3%) is followed by a non-significant decline of 17% between ICE-GB and FCSE. Usage in Broadcast Interviews & Discussions and Spontaneous Commentary shows a different pattern: A non-significant decline between LLC and ICE-GB (–22.8% and –34.8%) is followed by a more pronounced and significant increase between ICE-GB and FCSE (+91.1% and +133.8%).

Even if a growing readiness to use progressives with existence/relationship verbs is proposed on the basis of these numbers, it seems premature to conclude that progressive use with stative predicates is on the increase. This is due to the fact that not all progressive uses with existence/relationship verbs actually make reference to stative situations. Progressive *have* is most conventionalised in dynamic contexts such as 5.41. Similarly, progressive *be* (passive uses disregarded) is most frequent in connection with the *BE being adj.* pattern.

Overall, as Figure A.7 shows, progressive use with the two stative verb classes (stance and existence/relationship) remains fairly stable over time. Only in Parliamentary Language & Prepared Speech can a continuous upward trend be observed.

Excursion: Tracing the development of the *BE being adj.* pattern Before moving on to the analysis of the relative verbal frequency changes, the development of the *BE being adj.* pattern deserves a closer look. As has been mentioned above, the pattern frequently makes reference to momentary behaviour that is regarded/interpreted as deliberate or insincere (cf. chapter 3.3.2). Leech et al. (2009: 292) found tentative evidence for increasing usage in the Brown family corpora. This is backed by findings of Levin (2013), who reports increasing frequencies in 20th-century AmE. Given the fact that several authors regard the *BE being adj.* pattern as a special use of the interpretative progressive (Ljung 1980: 69ff.; Levin 2013: 193; Žegarac 1993: 214f.), its development seems particularly interesting.

In DCPSE and FCSE, the pattern reaches a frequency of altogether 55 tokens – too low for a sound diachronic analysis. If the genre distinction is disregarded and the normalised frequencies of LLC, ICE-GB and FCSE are compared, no clear trend emerges: In LLC, the pattern occurs 16 times, which amounts to a frequency of 4.3 tokens/htw. In ICE-GB, 20 tokens result in a normalised frequency of 5.8 tokens/htw. Finally, 19 tokens are found in FCSE, which corresponds to 4.1 tokens/htw. Thus, the pattern's frequency in DCPSE and FCSE is subject to a certain degree of fluctuation but exhibits no consistent upward development.

Due to the pattern's low raw frequency, it was decided to make use of two further corpora, BNC and Spoken BNC2014, which comprise several million words of spoken BrE conversation and allow a diachronic comparison of material from the 1990s and today. As Levin (2013: 195) has shown, usage of *BE being adj.* is most advanced in conversation, making BNC and Spoken BNC2014 the ideal corpus resources. Using POS tags and regular expressions, both corpora were searched for most possible realisations of the *BE being adj.* pattern.[24]

The results – summarised in Table 5.4 – are only partly in accordance with the frequencies found in DCPSE and FCSE. In BNC-DS (i.e. the demographically sampled part of BNC), the pattern reaches a frequency of 29.5 tokens per million words (i.e. 2.95 tokens/htw). This is considerably less frequent than the level of use in ICE-GB (5.8 tokens/htw) and shows that the latter does not faithfully reflect the actual frequency in late 20th-century spoken English. In Spoken BNC2014, however, the pattern occurs with a frequency of 43.34 tokens per million words (i.e. 4.34 tokens/htw), which comes very close to the FCSE frequency (4.1 tokens/htw). Given the huge amount of transcribed speech in the two BNC corpora, this statistically significant upward trend (+14 tokens pmw) faithfully indicates increasing usage of the *BE being adj* pattern.

Chapter 6 will study further lexical-grammatical progressive patterns that have acquired specific pragmatic constraints, analysing their development between the 1990s and today.

[24] Search term for BNC-DS: [pos="VBB|VBD|VBI|VBN|VBZ"] [word="not"]? [pos="AV0"]?[word="being"] [pos="AV0"]? [pos="AJ0|AJC|AJS"]; search term for Spoken BNC2014: [pos="VB0|VBDR|VBDZ|VBI|VBM|VBN|VBR|VBZ"] [word="not"]? [pos="RG|RGR|RGT|RR"]? [word="being"] [pos="RG|RGR|RGT|RR"]? [pos="JJ"].

	freq pmw[1]	word number
BNC94	29.51 (148)	5,014,655
BNC14	43.34 (495)	11,422,617
	***+46.9%	

Table 5.4.: *BE being adj.* in BNC-DS (conversation) (BNC94) and Spoken BNC2014 (BNC14). [1]Numbers in brackets indicate raw frequencies.

Verb classes: Relative frequency development

So far, the results of the verbal analysis are in many respects in line with Hypothesis 2. In those genres in which progressive use increases, usage with activity/event or – more generally – dynamic verbs contributes most to the change (in absolute terms). This means that further expansion of prototypical core uses is indeed driving the frequency increase.

However, also progressive use with the intermediate verbs – specifically cognition and perception/sensation verbs – has been identified as a noticeable contributor to increasing progressive use. In all genres, incl. Face-to-face Conversation, FCSE frequencies are much higher than those found in LLC. The finding is in line with Levin (2013: 201ff.), who reports increasing progressive use with mental verbs in 20th-century AmE. While most progressive uses with intermediate verbs do not qualify as real innovations, their increasing frequency is nonetheless an innovative feature, since it entails increased use in contexts in which the progressive is an optional but no obligatory grammatical encoding option.

In order to check whether the observed developments per htw also result in a significantly changed frequency distribution of the different verb classes relative to each other, let us consider the mosaic plots in Figure 5.7.

In all four plots, the eight different verb classes are conflated according to the dynamic/stative distinction that has been introduced above.[25] This broad classification indicates in general terms how the constructional network of the progressive has changed with regard to its verbal/semantic preferences.

The overall variation observed for Face-to-face Conversation is statistically significant ($p<0.05$). Interestingly, highest residuals can be observed for intermediate verbs. While their use is significantly less frequent than expected in LLC (–2.4), it is significantly overrepresented in FCSE (+2.1). In ICE-GB, basically no deviation can be observed. This means that intermediate verbs have developed from significant under- to significant overrepresentation between mid-20th-century English and today. Their share among progressive verbs in Face-to-face Conversation has continuously increased. Counter to this positive development, dynamic verbs show the reverse trend: They have a positive residual in LLC, a slightly negative one in ICE-GB, and a slightly more negative one in FCSE, meaning that their LLC uses are overrepresented in the data, while their ICE-

[25]The individual development of all eight verb classes is depicted in Figure A.8.

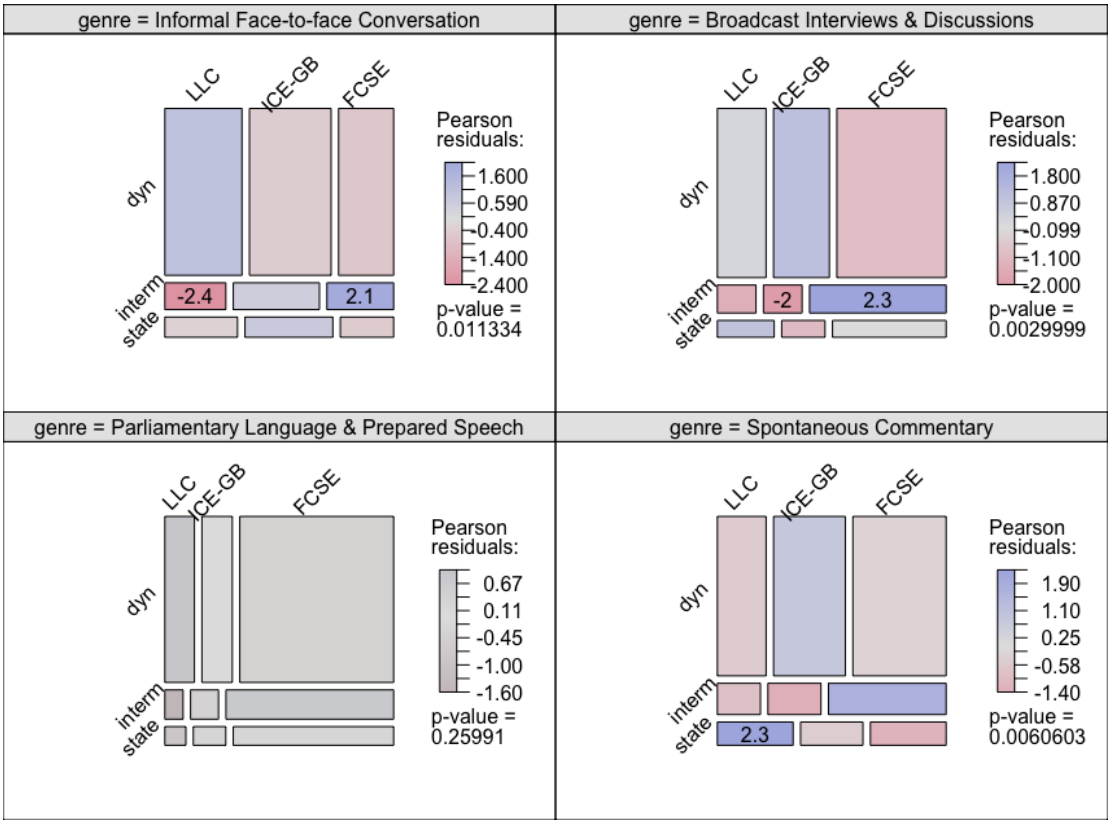

Figure 5.7.: Verb class by corpus and genre (active progressives only). dyn = activity/event + process/aspectual + communication, interm = causative + cognition/emotion/attitude + perception/sensation, state = existence/relationship + stance. Numbers and shading indicate Pearson Residuals.

GB and FCSE uses are underrepresented. Importantly, though, none of the respective residuals lies above/below +/–2, meaning that neither the overrepresentation in LLC nor the underrepresentation in ICE-GB and FCSE is statistically significant. State verbs show little variation in all three subcorpora. Very small non-significant negative residuals for LLC and FCSE (light red shading) stand against a non-significant positive residual of ICE-GB uses.

In Broadcast Interviews & Discussions, a similar picture emerges: Again, it is intermediate verbs that show significant variation. LLC and ICE-GB progressive uses of these verbs are less frequent than expected. However, only the ICE-GB uses get a significant residual (–2). Counter to this, FCSE progressive uses with intermediate verbs are significantly more frequent than expected (+2.3). The high frequency of intermediate verbs in FCSE also explains the (non-significant) negative residual for dynamic progressives in the same subcorpus. Co-occurrence with dynamic verbs is less typical of progressive use

in FCSE than it was in ICE-GB. Altogether, though, dynamic verbs show no dramatic degree of variation, none of the residuals reaching a level above/below +/–2. This is also the case for stative verbs, which exhibit a small degree of non-significant fluctuation.

The last genre that exhibits a statistically significant degree of overall variation is Spontaneous Commentary ($p = 0.006$). Here, state verbs show a negative development. Their LLC uses are significantly more frequent than expected (+2.3), while their ICE-GB uses come very close to the expected frequency. Finally, their FCSE uses are noticeably but non-significantly less frequent than expected. Counter to this trend, progressive uses with intermediate verbs are non-significantly overrepresented in FCSE and non-significantly underrepresented in LLC and ICE-GB, which is in accordance with their behaviour in Face-to-face Conversation and Broadcast Interviews & Discussions. Only very low positive and negative residuals are found for dynamic verbs. Their use with the progressive is slightly more typical in ICE-GB than in LLC and FCSE.

Basically no variation is present in the Parliamentary Language & Prepared Speech data. The overall p-value is very high ($p = 0.259$), far from reaching significance. For neither of the three verb classes can a distributional shift be observed. In the light of the fact that two significant normalised frequency changes have been observed for this genre (significant increase of activity/event and perception/sensation verbs), this is an interesting result. The mosaic plot shows that the relative share of the different verb classes has remained more or less stable, meaning that none of the individual frequency developments per htw was strong enough to change the relative verbal distribution in a noticeable way. The frequency changes per htw seem to counterbalance each other.

Summing up these observations, four insights come to the fore: First, in three of four genres, progressive use with intermediate verbs is more typical of the FCSE than of the DCPSE data. In Face-to-face Conversation and Broadcast Interviews & Discussions this overrepresentation reaches a significant level. In Spontaneous Commentary it is clearly present as well, but does not reach significance. Only in Parliamentary Language & Prepared Speech no such trend emerges. Thus, intermediate verbs have become significantly more typical of progressive use in the two conversational genres, Face-to-face Conversation and Broadcast Interviews & Discussions. In the less conversational genre of Spontaneous Commentary, the trend is visible but less pronounced. Finally, the lack of a trend in Parliamentary Language & Prepared Speech might be due to the supposedly conservative nature of the genre.

The second observation concerns progressive use with dynamic verbs. In none of the four genres is it overrepresented in the FCSE data. In the light of the pronounced frequency increase per htw in three of four genres, this is a noteworthy result. It shows that in all those genres in which normalised frequencies have increased, this increase has not resulted in a significantly increased share among all progressive uses. Relative to their previous level of use, progressives with intermediate verbs have increased in an even more pronounced way than progressives with dynamic predicates.

Third, the most conservative genre, Parliamentary Language & Prepared Speech, shows an exceptionally stable verbal paradigm. While significant distributional changes per htw can be observed, the distribution of verb classes relative to each other remains virtually unchanged over time.

Finally, progressive use in combination with verbs qualifying as stative is no more typical today than it was several decades ago. On the contrary, in one genre, Spontaneous Commentary, the reverse trend can be observed.

Individual verbs

Up to this point, the analysis of the progressive's verbal paradigm has remained on a relatively abstract level. Based on eight different verb classes and the dynamic/stative distinction, general trends have been revealed. Specific developments on individual verb level, however, have not been addressed. Consequently, the following section will move the focus from semantic classes to individual verbs, illuminating recent lexical-semantic developments of the progressive. This way, two major questions will be addressed: First, how does the development on verb class level "translate" to the level of individual verbs? Second, can significant developments on individual verb level be detected that are too small to play out on the level of semantic class and that have gone unnoticed so far?

Using *multiple distinctive collexeme analysis* (MDCA) (cf. Gries and Stefanowitsch 2004; Stefanowitsch 2013: 295ff.; Hilpert 2012), it will be possible to identify exactly those verbs that are distinctive of progressive use in each subcorpus and to reveal how the construction's verbal choices differ over time. As has been explained in chapter 4.4.3, MDCA is geared towards the detection of differences and will find verbs whose use is maximally distinct in the three corpora. Put simply, it will reveal which verbs are typical of which period (these verbs are called *distinctive collexemes*).

One of the advantages of MDCA is that it does not automatically "reward" high token frequency, meaning that also low-frequency verbs can potentially be identified as distinctive collexemes. This way, it is possible to reveal developments that would otherwise be missed.

A difficult question, however, regards the partitioning of the data. So far, all developments have been analysed along the lines of the four genre categories Face-to-face Conversation, Broadcast Interviews & Discussions, Parliamentary Language & Prepared Speech and Spontaneous Commentary. With regard to the comparatively small genre sizes, though, an analysis on individual verb level does not come without problems. Ultimately, it will lead to very low overall numbers for all but the most frequent verbs, diminishing the explanatory power of the analysis. Given these caveat, it was decided to distinguish only between Face-to-face Conversation on the one hand, and all other genres on the other hand.[26] This binary division is justified by the fact that – first – Face-to-

[26] MDCA does not require the data for each period to come from corpora of equal size. Thus, the different word count of Face-to-face Conversation in the three subcorpora is unproblematic. With regard to the conflation of the other three genres, however, one has to be aware that in the conflated genres of LLC and ICE-GB on the one hand, and of FCSE on the other hand, the individual genres are not equally represented (i.e. have a different ratio), which is due to differing word counts. Therefore, it was decided to construct a balanced version of the three subcorpora in which the size of each genre is equal in LLC, ICE-GB and FCSE. Both the original corpus data and the balanced version were fed into MDCA and the results were compared. Interestingly, they were very similar, exhibiting only minor differences. Most collexemes that turned out as distinctive for one subcorpus in the original data set also did so in the balanced version. Only their rank order was slightly changed. Given this

face Conversation has a substantial word count in all subcorpora and that – second – its overall frequency development differs from the one observed for the other genres. While a continuous frequency increase has occurred in Broadcast Interviews & Discussions, Parliamentary Language & Prepared Speech and Spontaneous Commentary, no such trend has been observed for Face-to-face Conversation.[27]

Let us first turn to the development in Face-to-face Conversation. Table 5.5 lists the distinctive collexemes of each subcorpus. It indicates the observed and expected frequency of each verb as well as the collostructional strength (i.e. the degree of attraction) and the significance level. It can immediately be seen that the observed frequencies of the distinctive collexemes differ substantially. While some have comparatively high token counts, others have counts as low as three. Nonetheless, all listed collexemes reach statistical significance (with the exception of those written in brackets).

LLC				ICE-GB				FCSE			
Verb	Obs	Exp	CollStr	Verb	Obs	Exp	CollStr	Verb	Obs	Exp	CollStr
teach	16	7.68	3.47***	come	65	48.80	2.64**	struggle	10	2.90	4.87***
work	44	29.33	3.11***	read	17	9.68	2.54**	chat	7	1.84	4.05***
check	4	1.40	1.83*	tape	6	2.32	2.47**	watch	11	4.48	2.99**
get up	4	1.40	1.83*	wear	13	6.97	2.38**	cry	6	1.84	2.74**
prepare	4	1.40	1.83*	hang	5	1.94	2.06**	miss	6	1.84	2.74**
complain	5	2.10	1.65*	save	5	1.94	2.06**	happen	16	8.17	2.63**
get	61	49.94	1.48*	look	46	34.47	2.06**	hold	6	2.11	2.25**
live	19	13.27	1.40*	use	17	10.46	2.03**	joke	8	3.43	2.10**
dig	3	1.05	1.37*	increase	6	2.71	1.80*	learn	7	2.90	1.99*
fall	3	1.05	1.37*	record	4	1.55	1.65*	study	5	1.84	1.79*
guard	3	1.05	1.37*	travel	4	1.55	1.65*	hairdress	3	0.79	1.74*
press	3	1.05	1.37*	sing	8	4.26	1.62*	cut	6	2.64	1.59*
rain	3	1.05	1.37*	wonder	13	8.13	1.57*	join	4	1.58	1.34*
set	3	1.05	1.37*	find	5	2.32	1.45*	worry	4	1.58	1.34*
treat	3	1.05	1.37*	describe	6	3.10	1.37*	(think)	35	27.41	1.22
				show	7	3.87	1.34*				

Table 5.5.: Distinctive verbal collexemes of the progressive (active) in Informal Face-to-face Conversation.

The distinctive collexemes of LLC Face-to-face Conversation are clearly dominated by dynamic verbs. Most of them are activity/event verbs (e.g. *teach, work, check*). With *complain*, one communication verb appears in the list. Only one verb, *live*, qualifies as stative. This dominance of activity/event / dynamic verbs is in line with the results of

similarity of results, it was decided to use the original corpus data for the present study. This comes with the advantage of a higher word count, resulting in higher observed frequencies of the individual verbs.

[27] Furthermore, the biggest difference in the development of one verb class (per htw) also exists between Face-to-face Conversation and the remaining genres: Progressive use with activity/event verbs shows no clear trend in Face-to-face Conversation but substantially increases in frequency in all other genres. As has been shown, however, this development looks differently if the verb classes' frequencies are analysed relative to each other (cf. Figure 5.7 & A.8).

the relative frequency analysis shown in Figure 5.7.[28] Interestingly, not one intermediate verb (i.e. cognition, perception/sensation, causative) occurs among the distinctive collexemes of LLC. Again, this is in line with the relative frequency development of the verb classes, where intermediate verbs are significantly underrepresented in LLC Face-to-face Conversation.

For ICE-GB Face-to-face Conversation, the picture is a bit more varied. While dynamic verbs still dominate and are highest ranked (*come, read, tape*), two state verbs (*wear, hang*) and three intermediate verbs (*look, wonder, find*) appear in the list. The occurrence of intermediate verbs reflects the fact that this class is no longer underrepresented in the respective mosaic plot in Figure 5.7.

Dynamic verbs also dominate the picture in FCSE (e.g. *struggle, happen, cut*). Only one state verb can be identified (*hold*). Intermediate verbs are *miss*, *worry* as well as *think*, which just about fails to reach statistical significance (but has a high observed frequency). Furthermore, *watch, learn* and *study* refer to perceptual or cognitive activities. It should also be noted that three of the dynamic verbs denote communicative activities (*chat, cry, joke*). Thus, it seems justified to argue that the distinctive collexemes of FCSE Face-to-face Conversation constitute the most diverse of the three sets. Compared to that, the LLC collexemes form a very homogeneous group. This finding is generally in line with the previous results of the verbal analysis, which suggest an expansion of the progressive's verbal range. Importantly, clear-cut state verbs do not seem to play a role in this development (in quantitative terms), showing no increasing use.

Before moving on to the MDCA results of the other genres, I would like to adopt a more qualitative perspective and focus on certain distinctive collexemes whose use with the progressive is particularly interesting. In order to arrive at such a selection, I went through all distinctive collexemes of each subcorpus and studied their progressive concordances. Naturally, in such a procedure collexemes with low observed frequencies had to be ignored, since a concordance of – say – three tokens does not allow for generalisations. Broadly speaking, this qualitative analysis yielded two results. First – not all too surprising – it is the supposedly non-prototypical, intermediate or stative verbs whose progressive use is most interesting. Second, certain verbs' appearance among the list of distinctive collexemes seems to be connected to corpus-specific characteristics or to idiosyncratic use of individual speakers. Take *teach*, for example, the highest ranked collexeme in LLC Face-to-face Conversation. On the one hand, the MDCA results show that activity/event verbs are favoured in the LLC progressive data. On a more specific level, though, the high ranking of the activity/event verb *teach* seems to be based on the fact that most recordings of LLC Face-to-face Conversation were made in a university context (Francis 1992: 197), in which teaching is a frequently talked-about activity. This, in turn, is likely to boost the progressive frequency of *teach*. While such an observation does not render the overall picture/tendency of the MDCA implausible, it

[28] In this respect, it is important to note that MDCA highlights asymmetries in relative frequencies. Consequently, it is the relative frequency development of verb classes against which the results should primarily be compared.

nonetheless shows that each distinctive collexeme needs to be studied carefully before (far-reaching) generalisations can be made. Apart from aspects of corps design, these considerations also concern idiosyncratic use of certain verbs by specific speakers.

Taking into account these considerations, it was decided that *live*, *look*, *wonder* and *think* should be analysed in more detail.

Live is among the distinctive collexemes of LLC Face-to-face Conversation and – with a frequency of 19 tokens – occurs about 1.5 times as frequently as expected (13 tokens). It is the only non-dynamic, i.e. stative, verb among the LLC collexemes, which makes it particularly interesting. With regard to its progressive use, the notion of *temporariness* is important. Since *live* refers to stative situations, the other two core aspectual properties of progressive meaning, *dynamicity* and *incompletion*, do not apply. When used with the progressive, one would expect uses such as the one in example 5.45, where *live* refers to a clearly temporary situation and focuses on the here and now (made explicit by the temporal adverb). While still not conveying dynamicity or incompletion, such uses are at least partially in accordance with the progressive's aspectual meaning.

(5.45) *[...] one of the girls in this flat* ***I'm living*** *in now was saying last night that [...].* (DCPSE:DL-B24)

(5.46) *But it isn't really suited to people who (,) who* ***are living*** *separate lives really.* (DCPSE:DL-B19)

On the other hand, there are uses such as the one in example 5.46 where no temporary element can be discerned and the progressive refers to an observation that is regarded as generally valid. As pointed out in chapter 3.4.2, Römer (2005) found comparatively frequent progressive use in exactly such contexts, which seem to defy any other clear-cut functional classification. While it is the simple aspect that appears to be the default option, progressive use is nonetheless acceptable, supporting the argument that has been made in connection with example 5.38.

In addition to that, progressive *live* can be used in contexts that do not allow a clear classification regarding temporariness and that are somehow vague. Interestingly, only four of the 19 progressive uses with *live* refer to clear temporary situations. Ten tokens are indeterminate, defying an unambiguous temporary/non-temporary classification, while five tokens refer to situations that can clearly be considered non-temporary. This means that the majority of progressive uses with *live* in LLC Face-to-face Conversation do not clearly exhibit any of the three central semantic characteristics of progressive aspectuality (progressivity/ongoingness, limited duration/temporariness, incompletion). As argued above (in the section on stance verbs' development per htw), this apparent lack of clear semantic constraints seems to be the major reason for the comparatively frequent use of progressive *live*. While it can be assumed that it was initially used to refer to temporary situations only, these uses have most likely paved the way for ambiguous as well as non-temporary uses. Importantly, such uses were already established in mid-20th-century spoken English and can be seen as a harbinger of further diversification of the construction's lexical-semantic range.

Among the distinctive collexemes of ICE-GB Face-to-face Conversation is *look*, a perception/sensation verb that can both be used dynamically and statively (cf. chapter 5.2.2). While its expected frequency lies at ca. 35 tokens, it occurs with an observed frequency of 46 tokens, making it significantly overrepresented. Analysing the individual occurrences, I distinguished between dynamic uses (example 5.29), stative uses (example 5.30) and metaphorical uses (*look forward*), referring to a an emotional state of mind (cf. example 5.47).

(5.47) *So* ***I'm not looking forward*** *to this rugby match.* (DCPSE:DI-B31)

Of the 46 tokens, 33 (about 70%) express dynamic meaning. The remaining 30% constitute – in equal shares – stative and metaphorical uses. This dominance of dynamic contexts is in line with the progressive's semantics. However, about one third of all uses do not refer to dynamic situations, showing that progressive *look* is also established in semantic contexts that can be considered non-prototypical from an aspectual point of view. As has been explained in the context of the progressive's relative frequency development with dynamic, intermediate and stative verbs, progressive use with intermediate verbs has developed from under- to overrepresentation in Face-to-face Conversation (cf. Figure 5.7). Frequent and semantically diverse use of progressive *look* in the ICE-GB data shows how this development plays out on individual verb level.

In chapter 3.4.4, it was argued that the progressive has to be analysed not only with regard to its morphosyntactic and semantic preferences but also as a lexical-grammatical phenomenon. While lexical-grammatical progressive patterns will be the focus of chapter 6, closer examination of another verb, *wonder*, already highlights the importance of (more or less variable) patterns of use. In ICE-GB Face-to-face Conversation, progressive *wonder* occurs 13 times (expected frequency = 8). The majority of uses (8 tokens) are part of the pattern *I was (just) wondering* (example 5.48). Two tokens form the present tense use *I'm just wondering* (example 5.49), and one token takes the past tense but a plural subject (*we were wondering*) (example 5.50). The remaining two tokens are somewhat different and cannot be said to follow any pattern. Importantly, though, none of the 13 tokens occurs with a third person subject, meaning that progressive wonder appears to be a feature of (inter-)subjective style. In many cases, its uses fulfil a discourse-pragmatic hedging function, rendering the following statement less definite (as for example in 5.48). Semantically, progressive *wonder* is not easily classified as dynamic or stative. While one could argue that a dynamic element is present in all 13 tokens (*wondering* in the sense of actively reflecting), one could also make the case for a stative interpretation, in which progressive *wonder* refers to a temporary state of mind. In almost all cases, it is very hard to decide which of the two meanings is present/dominates. Arguably, the majority of tokens is best described as indeterminate. From a semantic point of view, which does not take into account pragmatics, this would mean that progressive encoding of *wonder* is possible but not obligatory.

(5.48) *I* ***was wondering*** *when you were going to do that.* (DCPSE:DI-B49)

(5.49) ***I'm just wondering*** *if this is actually picking anything up.* (DCPSE:DI-B04)

(5.50) *Well we* ***were wondering*** *about that.* (DCPSE:DI-B58)

The final collexeme to be considered here is *think*. Progressive *think* is overrepresened in FCSE Face-to-face Conversation (35 observed vs. ca. 27 expected tokens), reflected in a collostructional strength of 1.22 – almost reaching statistical significance. It belongs to the class of cognition verbs (qualifying as intermediate on a dynamic/stative scale), whose use in connection with the progressive has increased between LLC and FCSE both per htw and relative to the other verb classes. As examples 5.25 and 5.26 show, progressive think can be used dynamically as well as statively. Consequently, similar to the analysis of *look*, all 35 tokens of progressive *think* were classified as to whether they refer to dynamic or stative situations. Interestingly, less than half of the cases (14 tokens, 40%) qualify as clearly dynamic. The remaining 60% (21 tokens) are either clearly stative (2 tokens, 6%) or defy a clear-cut categorisation and are best described as indeterminate cases (19 tokens, 54%). This means the majority of uses of progressive *think* is at best vaguely in agreement with the progressive's core semantics. Many tokens qualify as general imperfective uses, where the situation is presented from within but does not necessarily convey a dynamic element.

Apart from these semantic considerations, a second observation concerns the choice of subject: Most uses (23 tokens, 66%) occur in connection with a first person subject (mostly singular). Of the remaining cases, all but three tokens combine with a second person subject. This shows that – similar to *wonder* – use of progressive *think* is mainly a feature of involved, (inter-)personal style.

Having addressed Face-to-face Conversation, the following paragraphs will focus on the developments in the other genres. Table 5.6 lists the distinctive collexemes of each subcorpus. 17 of 21 LLC collexemes are dynamic verbs. Five of them denote communicative activities. While only one intermediate verb, *understand* (cognition), can be found, three verbs belong to the class of stative verbs (*live* and *line* (both stance) and *earn* (existence/relationship)). On this general semantic level, the list of ICE-GB collexemes looks quite similar: 11 of 13 verb types are dynamic (including two communication verbs), and two stative (*stand* and *wait*) (both stance). No verb belonging to one of the three intermediate verb classes (causative, cognition, perception/sensation) can be found, however. Counter to that, five intermediate verbs occur among the distinctive collexmes of the FCSE data (*expect, hope* (cognition), *feel, look* (perception/sensation), *help* (causative)). Two further intermediate verbs (*allow* (causative) and *intend* (cognition)) are also distinctive of the FCSE data but fail to reach statistical significance. In addition to that, *have* is among the distinctive collexemes together with *face*, the second stative verb on the list. The remaining ten verbs all refer to activities, events, processes or developments. Interestingly, no communication verb is among them.

As this overview shows, it is again the collexemes of the FCSE data that form the most diverse group of verbs. While the distinctive collexemes of LLC and ICE-GB are dominated by verbs referring to dynamic situations, plus the occasional stative verb, almost a third of the FCSE collexemes are made up by cognition, perception/sensation and causative verbs. Furthermore, *have* – a verb often considered an anti-progressive verb (cf. Stefanowitsch and Gries 2003: 230ff.) – is also significantly more frequent

LLC				ICE-GB				FCSE			
Verb	Obs	Exp	CollStr	Verb	Obs	Exp	CollStr	Verb	Obs	Exp	CollStr
deal	10	2.75	4.41***	march	5	1.37	2.52**	do	142	119.12	3.15***
come	39	23.75	3.15***	come	40	27.63	2.19**	expect	10	5.75	2.40**
express	5	1.18	2.84**	pull	5	1.60	2.07**	provide	13	8.06	2.31**
teach	3	0.59	2.12**	begin	12	6.16	1.98*	work	65	52.94	2.18**
approach	6	2.16	1.97*	chase	4	1.14	1.95*	help	15	9.78	2.14**
live	10	4.91	1.82*	damage	3	0.68	1.92*	have	53	42.58	2.05**
emerge	4	1.18	1.80*	quote	3	0.68	1.92*	look	104	89.77	1.91*
travel	4	1.18	1.80*	seek	9	4.34	1.79*	fail	7	4.03	1.68*
earn	5	1.77	1.74*	stand	12	6.85	1.58*	offer	7	4.03	1.68*
begin	10	5.30	1.57*	receive	4	1.37	1.56*	feel	17	12.08	1.65*
refer	5	1.96	1.52*	wait	11	6.16	1.54*	fall	12	8.06	1.58*
discuss	4	1.37	1.51*	write	5	2.05	1.47*	hope	12	8.06	1.58*
advance	2	0.39	1.41*	fly	3	0.91	1.40*	deliver	9	5.75	1.48*
bowl	2	0.39	1.41*					lose	6	3.45	1.44*
box	2	0.39	1.41*					put	23	17.84	1.37*
debate	2	0.39	1.41*					face	13	9.21	1.36*
handle	2	0.39	1.41*					fight	13	9.21	1.36*
line	2	0.39	1.41*					(allow)	8	5.18	1.28
overtake	2	0.39	1.41*					(intend)	5	2.88	1.20
understand	2	0.39	1.41*								
talk	36	27.08	1.41*								

Table 5.6.: Distinctive verbal collexemes of the progressive (active) in all genres other than Informal Face-to-face Conversation.

than expected. Thus, the FCSE collexemes are characterised by a higher number of supposedly non-prototypical progressive verbs than the collexemes of LLC and ICE-GB. Again, this is in line with the previous results of the verbal analysis, which suggest a diversification of the progressive's verbal paradigm, especially with regard to intermediate verbs.

In one aspect, however, LLC collexemes are more diverse than the ones of FCSE: While the former comprise five communication verbs (*express, refer, discuss, debate, talk*), not one appears among the latter. At first sight, this seems to be at odds with the results presented in chapter 5.2.2, which revealed (slightly) increasing use of communication verbs (between LLC and FCSE) in all genres except for Parliamentary Language & Prepared Speech). How can the distinctive collexemes of the conflated genre category seemingly suggest the opposite? The crucial point in this respect concerns the fact that MDCA compares frequencies relative to each other, while the figures presented in chapter 5.2.2 stand for themselves by presenting frequencies independently of each other per htw. Even though in absolute terms progressives with communication verbs have become more frequent, it was other verb classes – cognition and perception/sensation – that contributed more strongly to the progressive's increasing use, gaining a higher relative share among the different verb classes. While these "new" verbs were still infrequent in the LLC data, progressive use with communication verbs was – except for the genre of Spontaneous Commentary – already more frequent back then (above all in

Broadcast Interviews & Discussions, where it accounts for a high share of all progressives and where the portrayed development is especially evident). This is exactly the reason why several communication verbs occur among the LLC collexemes and several cognition and perception/sensation verbs among those of FCSE.

Regarding the analysis of individual verbs, it was decided to have a closer look at *live, talk, have, look* and *feel*.

As in Face-to-face Conversation, *live* is distinctive of LLC progressive use. Its observed frequency (10) is twice as high as its expected frequency (5). Of the ten tokens, only two refer to clearly temporary situations, while another two are hard to classify and are considered indeterminate. The remaining six uses refer to non-temporary situations, showing no obvious semantic overlap with progressive meaning (i.e. no notion of temporariness). This is in line with the observations on the use of progressive *live* in LLC Face-to-face Conversation, where temporary uses are also in the minority. Thus, *live* is distinctive of LLC progressive use both in Face-to-face Conversation and in the other genres, and it is most commonly used in non-temporary or indeterminate contexts. As has been observed for Face-to-face Conversation, it is frequently used in the context of generally valid statements. This notion of general validity is also pervasive in several uses of the conflated genre category, as the following two examples show:

(5.51) *Therefore (,) people in housing estates (,)* ***are (,) living*** *on the charity of their neighbours.* (DCPSE:DL-D01)

(5.52) *In (,) Elizabethan England („) the people who were most likely to be affected on the one hand („) were the people who* ***were living*** *(,) on relatively stable rents („).* (DCPSE:DL-J02)

In both examples, progressive encoding is clearly not obligatory and the simple aspect could just as well have been used. Thus, *live* represents a verbal context in which non-obligatory progressive encoding has gained a foothold several decades ago.

The second distinctive LLC collexeme that will be more closely analysed is *talk*. Its observed frequency (36) is about 33% higher than its expected one (27). A first examination of all tokens reveals that the vast majority – 31 of 36 tokens – were uttered in Broadcast Interviews & Discussions. Tokens stemming from Parliamentary Language & Prepared Speech and Spontaneous Commentary are much less frequent. Given the fact that the progressive is much more frequent with communication verbs in Broadcast Interviews & Discussions than in the other two genres (cf. Figure 5.6), this observation does not come as a surprise.

The functional analysis reveals semantically diverse use: 10 tokens express aspectual, progressive meaning (cf. example 5.53), 17 tokens are used in predominantly explanatory contexts in which they fulfil an interpretative function (5.54), and five tokens refer to recent communicative events (recentness progressive) (5.55). The meaning of the remaining four tokens could not be determined.

(5.53) *The four canons of Westminster (,) stand waiting to greet her („) and (,) Princess Anne* ***is talking*** *away to somebody at the moment (,).* (DCPSE:DL-F06)

(5.54) *That's to say when he says planning („) and the ordinary bloke on the Labour Party floor says planning (,) they're using the same words but they* ***'re talking*** *about different things.* (DCPSE:DL-D07)

(5.55) *I mean I think it's true you see even in (,) the kind of graduate seminar work I* ***was talking*** *about earlier on [...].* (DCPSE:DL-D07)

Interestingly, it is not the prototypical aspectual meaning that accounts for the majority of uses, but the subjective interpretative function, which – apparently – had already been well-established in mid-20th-century spoken BrE. This is in line with Smith's (2005: 92) claim that interpretative progressives and communication verbs frequently co-occur (cf. chapter 3.4.2, example 3.15) and with the observation made in chapter 5.2.2, which identified interpretative use of communication verbs as a discourse-pragmatic strategy in Broadcast Interviews & Discussions (cf. examples 5.22-23). As has been explained above, these interpretative uses – or, more precisely, the communication verbs used in these functional contexts – stick out as typical of the LLC data because other (non-aspectual) uses – e.g. progressive *have, expect, feel* or *hope* – were still on a lower level.

Let us now turn to progressive *have*. In the conflated genre category of FCSE, it occurs significantly more frequent (53 tokens) than expected (43), making it a distinctive collexeme. *Have* is clearly no typical progressive verb (in the analysis above, it was classified as an existence/relationship verb). As has been explained, progressive encodings of *have* are possible (cf. examples 5.41-43), even though they are very infrequent compared to simple aspect encodings. The only contexts in which progressive *have* seems conventionalised are dynamic ones as in example 5.41, where *have* refers to the active organisation and realisation of an event. Examples of progressive *have* that refer to (temporary) stative situations (example 5.42) or convey deontic meaning (example 5.43) can be found in the data – i.e. do not seem to be entirely unacceptable – but are certainly much less conventionalised.

In order to further illuminate the use of progressive *have*, all 53 tokens were semantically classified. In addition to the dynamic, stative and deontic uses, I included a further category, future uses, to be able to account for all tokens. The resulting picture is surprisingly diverse. Somehow unexpectedly, dynamic uses (example 5.56-58) account for only 19 of the 53 tokens (35.85%).

(5.56) *That's the tone in which we* ***should be having*** *this debate.* (FCSE_BI_10_38)

(5.57) *What discussions* ***is he having*** *with our allies in the five power defence arrangements?* (FCSE_PL_2_24)

(5.58) *They* ***'re having*** *a party at the start.* (FCSE_SC_10_181)

Stative uses account for almost the same amount of tokens (17, 32.08%), substantially adding to the overall token frequency of progressive *have*. While some of these uses sound more or less natural (5.59), others might be less acceptable (5.60). Finally, there are uses where it is hard to decide whether they constitute innovation, idiosyncratic

language use, or simply performance errors (5.61). Irrespective of the interpretation, though, it is certainly noteworthy that 17 of these stative uses can be found in such a small data set.

(5.59) *But people are still telling us that they **are having** problems and they are wondering why you've got an off shelf system [...].* (FCSE_BI_3_70)

(5.60) *These reductions **are having** a huge impact on the employment prospects of women in the public services.* (FCSE_PL_10_54)

(5.61) *[...] and he**'s having** a wonderful marathon career.* (FCSE_SC_10_16)

Furthermore, eleven deontic uses could be identified (examples 5.62-64), accounting for 20.75% of all tokens of progressive *have*. Like the stative uses, they can either be seen as potential innovations or simply as examples of idiosyncratic or even bad usage.

(5.62) *It **is having** to change in many ways because of the Eurozone crisis.* (FCSE_BI_15_140)

(5.63) *[...] more and more people **are having** to rent privately, with all the insecurity that that can bring.* (FCSE_PL_8_167)

(5.64) *He**'s having** to do that, because he hasn't got anybody in midfield to play into.* (FCSE_SC_2_164)

Taken together, stative and deontic uses of progressive *have* account for more than 50% of the 53 tokens in the FCSE data. Given the fact that only dynamic uses like the ones in examples 5.56-58 seem to be conventionalised, this picture might be indicative of an ongoing change in the use of progressive *have* in spoken British English. Apart from dynamic uses, stative and deontic uses might be spreading. Interestingly, though, this observation has been made for the conflated genre category, comprising Broadcast Interviews & Discussions, Parliamentary Language & Prepared Speech and Spontaneous Commentary. In Face-to-face Conversation, the most conversational and informal spoken genre, progressive *have* is distributed more evenly over time and was not identified as distinctive collexeme of the FCSE data.[29]

Finally, four uses of progressive *have* were classified as futurate progressives (cf. chapter 3.4.3).[30] Their use is highlighted in the following two examples:

(5.65) *And she'd heard like one bit of the show and another bit put two and two together and thought I **was having** a boob job.* (FCSE_BD_6_162)

[29] However, also in Face-to-face Conversation, progressive *have* is more frequent than expected in the FCSE data, and less frequent than expected in ICE-GB and LLC. The results fail to reach statistical significance, though.

[30] Two of the 53 tokens of progressive *have* defied unambiguous classification and could not be assigned to any of the four classes.

(5.66) *I mean it's mind-blowing enough for some people that I have a job and I drive, let alone that I've got married and I* ***'m having*** *a baby!* (FCSE_BD_9_203)

Apart from *have*, *look* is among the collexemes distinctive of progressive use in FCSE (expected freq = 90, observed freq = 104). While in Face-to-face Conversation it was distinctive of the ICE-GB data, it reaches this status only ca. 25 years later in the other genres. But what does the functional picture look like? How are dynamic (cf. example 5.29), stative (5.30) and metaphorical (5.47) uses represented? In ICE-GB Face-to-face Conversation, it was the dynamic uses that accounted for the vast majority of tokens (ca. 70%). This is also the case for the FCSE data of the conflated genre category. Here, even 86% of the 104 tokens of progressive *look* make reference to dynamic situations. Thus, the pattern is mostly used in semantic contexts in which it is firmly established. Non-asepctual, stative uses account for ca. 12%, reaching a similar level as in ICE-GB Face-to-face Conversation (15%). Finally, metaphorical uses (*look forward*), only account for 3% of all uses, being much less frequent than in ICE-GB Face-to-face Conversation (ca. 13%). This, however, is not all too surprising if one assumes that *look forward* is mostly used in personal, conversational contexts, in which the interlocutors talk about their feelings, plans, etc. Importantly, though, progressive *look* does not seem to be used more innovatively (i.e. more frequently referring to stative situations) in the conflated genre category of FCSE than in ICE-GB Face-to-face Conversation. The share of non-prototypical stative uses is very similar.

Another perception/sensation verb occurring among the distinctive collexemes of FCSE is *feel*. While its observed frequency (17) is much lower than the one of *look*, it is nonetheless used significantly more frequent than expected (12). *Feel* can be used with three different meanings: to sense (*I feel the age*), to think (*What do you feel about it?*) and to touch (*They were feeling their way along the wall*). While the first two meanings qualify as stative, the last one refers to a dynamic situation, meaning that it can easily take the progressive. Even though it is possible to use also stative *feel* with the progressive, such uses are optional and – arguably – not (yet) fully conventionalised. Surprisingly, though, only one of the 17 tokens potentially allows a dynamic reading (example 5.67). All other uses refer to stative situations (examples 5.68-69).

(5.67) *But in actual fact, he* ***seems to be uh feeling*** *his calf muscle (,) as he walked back to join his teammates.* (FCSE_SC_6_274)

(5.68) *[...] I* ***'m really feeling*** *confident about it.* (FCSE_BI_4_19)

(5.69) *Now, Philip, how* ***are you feeling*** *and what's on your 'To Do' list just a few days before the games start?* (FCSE_BD_3_139)

Both the progressive in sentence 5.68 and the one in 5.69 sounds natural and acceptable. However, the simple aspect would have worked just as well. As argued above, these examples constitute possible but non-obligatory contexts of progressive use. Given the fact that a) progressive use with perception/sensation verbs is generally increasing and that b) all but one of the 17 tokens of progressive *feel* refer to stative situations,

one can assume that such non-obligatory uses of the construction are increasing and – consequently – becoming more and more acceptable and routinised.

Going through the list of distinctive FCSE collexemes in Table 5.6, one finds further examples of verbs that allow the use of the progressive but can certainly not be regarded as prototypical choices (and which belong to one of the three intermediate verb classes): *expect, hope, allow** and *intend**.[31] Consider the following examples:

(5.70) *What sort of contest* ***are you expecting****, Scott?* (FCSE_SC_1_7)

(5.71) *But at the same time you know that's what I****'m hoping*** *the prime minister will achieve.* (FCSE_BI_22_29)

(5.72) *They are higher risk, but what they* ***are allowing*** *is for patients who maybe have been waiting for quite some time for a deceased donor transplant to give them another option for transplant.* (FCSE_PS_13_212)

(5.73) *You know, obviously as a campaign we****'re intending*** *to work closely with other organisations [...].* (FCSE_BD_4_130)

Similar to the stative uses of progressive *feel*, these progressives do not come across as wrong. Importantly, though, all of them also work with the simple aspect and are not grammatically required. In (formal) written language, one would almost certainly expect these examples to be encoded in the simple aspect.

The fact that the distinctive FCSE collexemes (both of Face-to-face Conversation and the other genres) are more diverse with respect to the occurrence of such verbs than the LLC collexemes supports the above-made claim that the verbal paradigm of the progressive construction is slowly widening. While state verbs do not seem to be developing in any significant way, cognition, perception/sensation and causative verbs – i.e. the intermediate classes – are diffusing into accepted use and, consequently, stand out as typical of the FCSE data. While this process might have started slightly earlier in Face-to-face Conversation, where three intermediate verbs can already be found among the ICE-GB collexemes, it is now also well-advanced in the other, more formal spoken genres. This interpretation is in line with the results shown in Figure 5.7, where the three intermediate verb classes show a frequency increase relative to state and dynamic verbs in three of four genres (only Parliamentary Language & Prepared Speech has a remarkably stable semantic paradigm).

This process, however, must not be interpreted as a slow replacement of prototypical, dynamic verb classes by less typical ones. As the analysis of the verb classes' development per htw showed (cf. Figure 5.6 & A.7), dynamic verbs (most importantly activity/event verbs) show a clear positive frequency trend in those genres in which progressive use is increasing. Also the relative frequency analysis did not reveal a statistically significant underrepresentation of dynamic verbs in the FSCE data for any of the four genres. However, since progressive uses with "new" verb classes are sneaking in, dynamic verbs are slightly less typical today than they used to be in the LLC and/or the ICE-GB

[31] The verbs marked with * narrowly miss statistical significance.

data. This is reflected in the fact that the distinctive collexemes of FCSE are made up by a more diverse set of verbs than those of LLC (and in the case of the conflated genre category also than those of ICE-GB). Clear state verbs play no important role in this development. They exhibit relatively little variation, regardless of whether they are measured per htw, relative to the other verb classes, or as distinctive collexemes. On individual verb level, however, it is progressive *have* that shows an interesting development in the three formal spoken genres, where it is used more frequently than expected and in contexts in which progressive use is neither fully conventionalised nor obligatory.

5.2.3. Conclusion

While chapter 5.1 focused on the progressive's overall frequency development in DCPSE and FCSE, the present subchapter has zoomed in on the development of two central aspects of progressive use: morphosyntactic context and verbal paradigm.[32] The starting point of the investigation was Hypothesis 2, which assumes the following:

Hypothesis 2 The frequency increase of the progressive in the 20th and 21st centuries is based on further expansion of prototypical core uses as well as on structural or functional innovations. Structural and functional innovations are expected to have less impact on the corpus data than the development of the prototypical core uses.

With regard to the progressive's morphosyntactic context, the hypothesis was largely confirmed. If a frequency increase occurred, it was the present progressive – the most typical context of use – that contributed most (per htw). Increasing use in other, less prototypical morphosyntactic contexts – most notably combinations with modal auxiliaries but also with *to*-infinitives – also contributed to the overall increase, but on a much lower level. Among the modal progressives, *will* – arguably used as matter-of-course future – is the strongest contributor. As an analysis of BNC and BNC2014 data revealed, the *BE being adj.* pattern has also increased by almost 50% since the early 1990s. Again, however, the contribution of such a non-prototypical progressive context to the overall development is a modest one.

The development (per htw) of the present progressive almost mirrors the overall frequency development of the construction, showing how dominant this morphosyntactic realisation actually is. Consequently, the progressive's observed frequency decrease in Face-to-face Conversation between ICE-GB and FCSE is linked to a decrease in present progressives. While this was not predicted by the hypothesis, it shows how closely aligned the two phenomena are.

[32] It is clear that further aspects of the construction's use exist which could potentially be studied. An obvious candidate is the progressive passive, which has received considerable attention in previous research (cf. chapter 3). However, it shows little variation compared to the progressive active (a non-significant frequency increase between LLC and ICE-GB, and stable usage between ICE-GB and FCSE) (cf. Figure A.2). Consequently, it was decided to refrain from a detailed analysis and to directly move on to the study of complex variable combinations (chapter 5.3).

Analysis of the relative frequency development – i.e. the frequency of each morphosyntactic context relative to the other contexts – revealed that none of the four genres is characterised by a significantly over- or underrepresented use of the present progressive in FCSE. For those genres in which an overall frequency increase of the progressive occurred – Broadcast Interviews & Discussions, Parliamentary Language & Prepared Speech and Spontaneous Commentary – this means that even though the present progressive has contributed most in absolute terms (i.e. additional tokens per htw), this contribution was still not pronounced enough (against the background of, for example, the development of modal progressives) to result in a significantly increased frequency relative to the other morphosyntactic contexts. Thus, one can argue that the present progressive is more entrenched/routinised in the language system today than 50 years ago, but not necessarily a more typical/central node in the progressive's constructional network.

In Face-to-face Conversation, however, where progressive frequency peaks in the 1990s, this peak is actually connected to a significant (relative) overrepresentation of present progressive use. The frequency increase between LLC and ICE-GB Face-to-face Conversation was not just linked to an increase in the present progressive per htw, but also to a significant relative frequency increase of this constructional variant.

Finally, it was observed that the relative frequencies of present and past progressive behave in opposite ways in all four genres. If the present progressive increases its relative share, the past progressive decreases, and vice versa. Apart from this, no consistent pattern of past progressive development could be made out.

The analysis of the progressive's verbal paradigm revealed that activity/event verbs are by far the most frequent semantic class. Similar to the present progressive, their frequency per htw corresponds relatively closely to the progressive's overall frequency development. Consequently, increase in progressive frequency goes along with increasing use of activity/event verbs, which is in line with Hypothesis 2. The frequency decrease between ICE-GB and FCSE Face-to-face Conversation is linked to a decrease in the use of activity/event verbs (per htw).

Another major observation regarding the progressive's semantic paradigm concerns cognition and perception/sensation verbs. Without exception in any of the genres, their frequency per htw is higher in FCSE than in LLC. Even though their contribution per htw to the overall frequency increase is lower than that of activity/event verbs, it should not be neglected (especially if the two verb classes are taken together). Percentage-wise the development is in most cases even more pronounced than that of activity/event verbs. This, again, is reflected in the relative frequency development, where intermediate verbs (i.e. those not clearly qualifying as dynamic or stative) have become more typical in three of four genres. The effect is significant in Face-to-face Conversation and Broadcast Interviews & Discussions, not however in Spontaneous Commentary. For Parliamentary Language & Prepared Speech, whose verbal paradigm is very stable in relative terms, no such trend can be observed, though.

Finally, the collexeme analysis revealed that the distinctive collexemes of FCSE (both in Face-to-face Conversation and the conflated genre category) are noticeably more diverse than those of LLC, comprising verbs such as *have*, *feel*, *expect*, etc.

All this evidence suggests that – apart from further expansion of prototypical progressive use with activity/event verbs (in those genres in which an overall frequency increase occurred) – the construction's semantic/verbal paradigm has widened in recent decades. While the progressive's semantic core does not seem to have changed in any significant way, less prototypical, often non-aspectual uses with cognition or perception/sensation verbs or with *have* have become more common. Thus, it is the peripheral parts of the construction's semantic network that are slowly changing.

5.3. Analysing the frequency development II: Variable combinations

This subchapter presents the last part of the analysis of the progressive's recent frequency development in DCPSE and FCSE, focusing on complex variable combinations.

For example, we have seen that the present progressive as well as progressives co-occurring with activity/event verbs exhibit a consistent frequency increase per htw in three of four genres. Thus, there are two prototypical contexts of use that are becoming more and more routinised. While this is certainly an interesting insight, it comes with a number of further questions: How do the two variables develop in combination? Are present progressives co-occurring with activity/event verbs increasing and, if so, to what extent? Is this most prototypical of all combinations becoming more typical still? If yes, in which genres?

Furthermore, it has been shown that progressive use with intermediate verb classes is increasing. But how is this increase interacting with different morphosyntactic choices in different genres?

In order to address these questions in a comprehensive way, we are in need of an analytic tool that can handle several variables at the same time and that proceeds in an exploratory fashion. As has been explained in chapter 4.4.4, one tool capable of performing such a task is *Hierarchical Configural Frequency Analysis* (HCFA) (Von Eye 1990; Von Eye 2002; Gries 2009: 240ff.; Hilpert 2013a: 55ff.). Making use of HCFA, we can expect to discover new and more fine-grained aspects of the progressive's use and development in spoken English.

5.3.1. Identifying types and anti-types: Hierarchical Configural Frequency Analysis

First, it had to be decided exactly which variables and variable levels were to be included in the analysis. On the one hand, a higher number of variables can potentially reveal more fine-grained insights than a smaller number. However, if too many variables are included, the results of the HCFA become extremely hard to interpret and statistically less robust.[33] In order to avoid such complications, the analysis focuses on subcorpus,

[33] Different from a single chi-square test, variable combinations are allowed to violate the assumption that expected frequencies always have to be larger than five. Instead, Von Eye (2002: 66) suggests that the squared overall number of observations, divided by the number of table cells, should be

genre, morphosyntactic context and verb class. Further variables such as voice or co-occurring adverbials are not taken into account.

Regarding variable levels, Hilpert explains that "[i]n order to keep the observed frequencies as large as possible, it is advisable to choose categorical variables that only have a small number of levels" (2013a: 61f.). In this respect, subcorpus (three levels) and genre (four levels) are unproblematic. Also the five different morphosyntactic levels are still acceptable. With regard to the eight different verb classes, however, matters are different. Apart from the high number of different levels, many of them have comparatively low token counts. Thus, it was decided to conflate some of the levels and to form broader categories. One option would have been to use the dynamic/intermediate/stative distinction. However, since HCFA wants to detect fine-grained constructional subschemas, this threefold division seemed slightly too crude. Eventually, I decided on a four level compromise that looks as follows: activity/event + process/aspectual verbs (level 1), communication verbs (level 2), cognition + perception/sensation + causative verbs (= intermediate verbs) (level 3), stance + existence/relationship verbs (= stative verbs) (level 4).

Another consideration concerns the different genre sizes. The fact that different genres are of different size does not pose a problem, since the expected frequencies are adjusted accordingly. What is problematic, though, is that also the same genres are not always of equal size in the different subcorpora. Since HCFA does not use normalised but raw frequencies, these differences in size between the same genres would complicate the interpretation of the results to an extreme degree. Thus, it was decided to use a balanced version of the corpus data, in which the same genres are of equal size in all three subcorpora. While this comes with the drawback of a lower number of progressive tokens ($N = 6266$), it was indispensable if a meaningful and valid interpretation of the results was to be ensured.

The HCFA results are presented in Table 5.7 (types) and 5.8 (anti-types). In each table, every row stands for a variable combination that is either significantly more frequent (Table 5.7) or less frequent (Table 5.8) than expected. The first four columns indicate the level of each of the four variables, while columns five and six specify the observed and expected frequency of the respective variable combination. Column seven indicates whether the observed frequency lies above (>) or below (<) the expected one. Finally, column eight indicates the significance level.

Let us first see if and how the supposedly most prototypical combination of verb class and morphosyntactic context – activity_process[34] & present[35] – is represented among the types and anti-types. The combination occurs three times among the types and one time among the anti-types, which indicates that it is subject to a substantial degree of variation. Among the types, it is part of the following variable combinations: (i)[36] *FCSE + Spontaneous Commentary + activity_process + present*, (ii) *FCSE + Parliamentary*

larger than ten (see also Hilpert 2013a: 61). This precondition is met in the present analysis.

[34] activity_process = activity/event & process/aspectual verbs.

[35] present = present progressive (active).

[36] Each type and anti-type is assigned a Roman numeral by which it will be referred to in the text. This numeral is independent of the rank order in Table 5.7 and 5.8.

Corpus	Genre	Verb	MSC	Obs	Exp	Obs/Exp	Sig.Level
FCSE	Spont.Comm.	intermediate	modal	12	1.91	>	***
FCSE	Spont.Comm.	activity_process	present	183	126.72	>	***
FCSE	F-F-Conv.	communication	past	82	47.55	>	***
FCSE	Parl.Lang.&Prep.Sp.	activity_process	present	126	82.37	>	***
FCSE	Broadcast Int.&Dis.	intermediate	present	75	43.24	>	**
FCSE	Broadcast Int.&Dis.	communication	perfect	16	4.48	>	**
ICE-GB	F-F-Conv.	activity_process	present	539	441.90	>	***
ICE-GB	F-F-Conv.	communication	past	72	45.97	>	*
LLC	F-F-Conv.	communication	past	73	36.39	>	***
LLC	Parl.Lang.&Prep.Sp.	activity_process	past	52	24.12	>	***
LLC	Broadcast Int.&Dis.	communication	present	70	43.24	>	*

Table 5.7.: HCFA types based on subcorpus, genre, semantic verb class, and morphosyntactic context (MSC) (active progressives only).

Corpus	Genre	Verb	MSC	Obs	Exp	Obs/Exp	Sig.Level
FCSE	F-F-Conv.	activity_process	present	359	457.07	<	***
FCSE	Broadcast Int.&Dis.	activity_process	past	46	79.50	<	**
FCSE	Spont.Comm.	communication	present	14	34.45	<	*
FCSE	F-F-Conv.	activity_process	perfect	16	36.25	<	*
ICE-GB	Spont.Comm.	communication	present	9	33.31	<	***
ICE-GB	Spont.Comm.	activity_process	past	18	46.87	<	***
ICE-GB	Broadcast Int.&Dis.	intermediate	present	20	41.80	<	*
LLC	Spont.Comm.	activity_process	past	5	37.11	<	***
LLC	Spont.Comm.	communication	present	4	26.37	<	***
LLC	F-F-Conv.	intermediate	present	39	72.79	<	**

Table 5.8.: HCFA anti-types based on subcorpus, genre, semantic verb class, and morphosyntactic context (MSC) (active progressives only).

Language & Prepared Speech + activity_process + present, and (iii) *ICE-GB + Face-to-face Conversation + activity_process + present*. Among the anti-types, it occurs in the pattern (iv) *FCSE + Face-to-face Conversation + activity_process + present*.

This is a very interesting result with regard to Hypothesis 2: It means that in three of four genres (Face-to-face Conversation, Parliamentary Language & Prepared Speech, Spontaneous Commentary), present progressives co-occurring with activity/event or process/aspectual verbs are overrepresented in the period in which progressive frequency is highest. For Face-to-face Conversation, this is the early 1990s (i.e. ICE-GB), for Parliamentary Language & Prepared Speech and Spontaneous Commentary this is today (i.e. FCSE). Only in connection with Broadcast Interviews & Discussions is the pattern never over- or underrepresented.

As anti-type (iv) tells us, the pattern is significantly less frequent than expected in FCSE Face-to-face Conversation, meaning that the recent decline of the progressive's

overall frequency in this genre is linked to a less frequent than expected use of the most prototypical realisation of the construction. While chapter 5.2 has already shown that the decline is linked to a decrease (per htw) in uses with activity/event verbs and with the present tense, the HCFA results – by revealing that the combination of the two variables (plus process/aspectual verbs) is significantly less frequent in Face-to-face Conversation today than in the 1990s – prove that these two developments are connected. The combination occurs with an observed frequency of just 359 tokens, compared to an expected frequency of ca. 457 tokens.

Activity/event and process/aspectual verbs also occur in other combinations that are significantly over- or underrepresented. Especially in connection with the past tense can considerable variation be observed. Among the types, past progressives plus activity_process verbs occur in the following combination: (v) *LLC + Parliamentary Language & Prepared Speech + activity_process + past*. Among the anti-types, they are found three times: (vi) *FCSE + Broadcast Interviews & Discussions + activity_process + past*, (vii) *ICE-GB + Spontaneous Commentary + activity_process + past*, and (viii) *LLC + Spontaneous Commentary + activity_process + past*.

Let us first consider the type: past progressives with activity/event and process/aspectual verbs occurring in LLC Parliamentary Language & Prepared Speech are significantly more frequent than expected. Interestingly, the overall progressive frequency of this period and genre is the lowest of all periods and genres (cf. Figure 5.1).

Counter to this, the activity_process + past pattern is part of anti-type (vi) *FCSE + Broadcast Interviews & Discussions + activity_process + past*, meaning that its use is less frequent than expected in FCSE Broadcast Interviews & Discussions, the period and genre with the highest overall progressive frequency (cf. Figure 5.1). In the light of this observation, one might want to argue that the activity_process + past pattern is typical of low and atypical of high levels of progressive use. However, this reasoning is not supported by anti-types (vii) and (viii), which reveal that the pattern is underrepresented in both LLC and ICE-GB Spontaneous Commentary, which both reach comparatively low overall progressive frequencies.

Again, as in chapter 5.2, analyses involving the use of the past progressive are difficult to interpret, meaning that no common trend that holds across genre boundaries can be identified. The most likely interpretation at this point seems to be that the development of the activity_process + past pattern somehow mirrors the development of the past progressive in the respective genres. Past progressive use is high in LLC Parliamentary Language & Prepared Speech (per htw and in relative terms), but low in FCSE Broadcast Interviews & Discussions (esp. in relative terms), and in LLC and ICE-GB Spontaneous Commentary (per htw and – for LLC – in relative terms) (cf. Figure 5.3 & 5.5).

Finally, there is one more anti-type in which activity/event and process/aspectual verbs occur: (ix) *FCSE + Face-to-face Conversation + activity_process + perfect*. While it is expected to occur ca. 36 times, its observed frequency is as low as 16 tokens. The combination does not seem to represent a pattern with a specific discourse-pragmatic or genre-specific function. However, its low frequency reveals that the comparatively infrequent use of activity_event verbs in FCSE Face-to-face Conversation interacts with the very infrequent use of the progressive perfect in same period and genre.

If we further examine the results, we discover that there are two types and two anti-types that involve intermediate verbs. In chapter 5.2, it has been argued that the progressive's verbal paradigm is slowly widening and that uses with intermediate verbs have become more frequent. Are these assumptions in accordance with the respective (anti-)types? The two types are composed as follows: (x) *FCSE + Spontaneous Commentary + intermediate + modal* and (xi) *FCSE + Broadcast Interviews & Discussions + intermediate + present*. The anti-types are (xii) *ICE-GB + Broadcast Interviews & Discussions + intermediate + present* and (xiii) *LLC + Face-to-face Conversation + intermediate + present*. Just focusing on the subcorpora of these (anti-)types, we see that combinations involving intermediate verbs are typical of present-day usage (FCSE) and underrepresented in the recent past (LLC and ICE-GB), which is generally in line with the results of chapter 5.2. Let us have a closer look at each variable combination.

Type (x) represents a very specific kind of progressive use. Intermediate verbs plus modal auxiliaries in FCSE Spontaneous Commentary. The observed frequency is not very high (12), but nonetheless about six times more frequent than expected. The use of this pattern is highlighted in the following examples:

(5.74) *She**'ll be definitely wanting** to prove a point, you know, she, her place is up for grabs here, and if, if she puts in a good performance then she might have a chance to start against France.* (FCSE_SC_4_35)

(5.75) *The man who **will be looking** for yellow is the rider who finished on second place in the stage yesterday, and he's a member of the Cannondale squad.* (FCSE_SC_3_137)

(5.76) *[...] I think that's what they**'ll be looking** to do in this ninety minutes with the four four two, banking up and making it hard for them to (,) be broken down [...].* (FCSE_SC_4_26)

Ten of twelve uses are formed with the modal *will* and qualify as matter-of-course futures. The vast majority of tokens (eleven) were uttered in the context of sport competitions and are used by the speakers to make predications about what is going to happen in the game, race, etc. *Look* dominates the picture (eight of twelve tokens), but is used with different meanings (mostly in a sense similar to *aim* (5.75-76) but also clearly dynamically in the sense of *search*). What is also noteworthy is the fact that ten of twelve tokens were uttered by just three different speakers – indicating that this specific progressive use is linked to idiosyncratic preferences. Nonetheless, the pattern shows an interaction of two trends that have been reported previously in the present work. First, increasing use of intermediate verbs and, second, increasing use of modal progressives formed with *will*. In Spontaneous Commentary, these two trends interact to a certain degree, leading to a sixfold overrepresentation of an otherwise very infrequent progressive context. Although this does not have a strong statistical impact on the progressive's frequency increase in this genre, it shows how the interaction of different developments can influence the progressive's occurrence in very specific contexts.

Unsurprisingly, progressives with intermediate verbs are more frequent in connection with the present tense. This combination is part of type (xi) and anti-type (xii) and (xiii).

Type (xi) reflects frequent progressive use of present tense intermediate verbs in FCSE Broadcast Interviews & Discussions. This variable combination occurs 75 times as opposed to ca. 43 expected tokens, which is a substantial overrepresentation. Exactly the same use is atypical of the same genre in ICE-GB (anti-type xii) (observed freq = 20, expected freq = 42). In chapter 5.2, Figure 5.7 showed that intermediate verbs are significantly overrepresented in FCSE Broadcast Interviews & Discussions and significantly underrepresented in the ICE-GB data of the same genre. This is in line with the reported types and and anti-types for ICE-GB and FCSE Broadcast Interviews & Discussions. However, due to HFCA, we are now able to say that the development of intermediate-verb progressives has mainly occurred in the present tense. Usage is illustrated in examples 5.68 and 5.69 as well as 5.71 and 5.73. As has already been argued above, progressive use in such contexts is optional and often (inter-)subjective, which is reflected in the use of first and second person pronouns.

As anti-type (xiii) shows, present tense intermediate verbs are underrepresented in LLC Face-to-face Conversation. Instead of ca. 73 expected tokens only 39 tokens occur in the data. Again, this reflects an interaction of two trends that have been observed in chapter 5.2: Comparatively low frequency of the present progressive (cf. Figure 5.3 & 5.5) and of intermediate verbs (cf. Figure 5.6 & 5.7) in LLC Face-to-face Conversation. These two trends jointly lead to an underrepresentation of present progressives with intermediate verbs.

While it has been argued in chapter 5.2 that the progressive is increasingly used in contexts in which it is optional (with intermediate verbs), it has also been claimed that stative verbs (stance and existence/relationship) show little variation on a level above the individual verb. The results of the HCFA further support this observation, since no type or anti-type was found that is formed with stative verbs. This means that progressive uses involving stative verbs show no significant variation in the corpus data.

Finally, there is one verb class that is part of five types and three anti-types: communication verbs. Three of the types are particularly interesting. These are (xiv) *FCSE + Face-to-face Conversation + communication + past*, (xv) *ICE-GB + Face-to-face Conversation + communication + past* and (xvi) *LLC + Face-to-face Conversation + communication + past*. Past progressive use with communication verbs is overrepresented in Face-to-face-Conversation in all three subcorpora.

As has been explained in chapter 5.2, the combination of past progressives with communication verbs often indicates the use of the recentness progressive, which refers to recent communicative happenings (cf. example 5.24) and often lacks clear aspectual meaning. Based on what I argued in connection with the development of the past progressive and of communication verbs in chapter 5.2, overrepresentation of the pattern could have been expected in FCSE Spontaneous Commentary. It is in this genre that both the past progressive and progressives with communication verbs increased. Cross-tabulation of the two phenomena indicated increasing use over time. However, as the HCFA results show, this development was not strong enough to produce a respective

type for FCSE Spontaneous Commentary. Instead, we see that the past progressive's co-occurrence with communication verbs is clearly a feature of Face-to-face Conversation. This overrepresentation is not part of a trend but has been stable for several decades.

In order to avoid misinterpretation of the data, it is imperative to check whether the tokens behind the *Face-to-face Conversation + communication + past* pattern do confirm the assumption that it actually represents uses of the recentness progressive. While certainly not all tokens in the three subcorpora make reference to recent communicative happenings, many examples can be found that are in line with the claim. The following sentences give a number of examples form LLC, ICE-GB and FCSE.

(5.77) *I **was just telling** uhm (,) Mike about my uh (,) this super Guinness beer I made.* (DCPSE:DL-B07)

(5.78) *He **was asking** me about this today.* (DCPSE:DL-B23)

(5.79) *No I **was just saying** who d who who did you mean.* (DCPSE:DI-B63)

(5.80) *So uhm as I **was saying** („) there might be an amendment.* (DCPSE:DI-B54)

(5.81) *And I think, like, we go back to earlier, when I **was saying** to you (,) about this, you know, you can go anywhere with your career, you know.* (FCSE_FFC_19_190)

(5.82) *[...] what we **were saying** earlier was that our kind of friend, like circumstances have changed massively.* (FCSE_FFC_3_1538)

Admittedly, the recentness progressive is a fairly vague functional label and every attempt to identify all tokens is possibly futile since the classification procedure includes a very high degree of subjectivity and uncertainty. Nonetheless, the present results indicate that this specific non-aspectual progressive use is a feature of Face-to-face Conversation. This view is exactly in line with Leech's observation that there is "one common application of the Past Progressive in conversation" where it "refers [...] to fairly recent communicative happenings" (Leech 2004: 32).

Another type that comprises communication verbs is (xvii) *FCSE + Broadcast Interviews & Discussions + communication + perfect*. With 16 tokens, it is not very frequent; however, its expected frequency is much lower still (ca. 4.5). Consider the following examples:

(5.83) *You know I**'ve been suggesting** for a long time you try and create some safe havens for you know the Syrian opposition to operate from.* (FCSE_BI_16_103)

(5.84) *And I**'ve been talking** rubbish for nearly 13 years now.* (FCSE_BD_1_173)

(5.85) *[...] so we**'ve just been talking** about how important the hospital, the stadium and the whole Stoke Mandeville legend has been.* (FCSE_BD_3_163)

(5.86) *The Prime Minister's been very, very focused on the situation we**'ve just been describing**.* (FCSE_BI_15_113)

While the progressives in 5.83 and 5.84 make reference to a repeated communicative act that extends over a long period leading to the present, the ones in 5.85 and 5.86 refer to recent communicative activities that extend only over a short period of time. They are functionally similar to the recentness progressive: If one compares the progressive in 5.77 to the one in 5.85, a change from the present perfect to the past tense (and vice versa) does not seem to make much of a semantic and pragmatic difference. Since these uses account for roughly half of the 16 tokens of the pattern, its overrepresentation in Broadcast Interviews & Discussions might be linked to their increased occurrence.

Finally, there is one more type involving communication verbs: (xviii) *LLC + Broadcast Interviews & Discussions + communication + present*. It is considerably more frequent (70 tokens) than expected (ca. 43 tokens), and many of its realisations represent argumentative progressive uses that are based on the construction's interpretative meaning.

(5.87) *What they**'re saying** is (,) if you're earning twenty pounds a week you must get out (,).* (DCPSE:DL-D01)

(5.88) *(„) I**'m not suggesting** that you should feel guilty [...].* (DCPSE:DL-D08)

(5.89) ***Are you saying** that Mr Chapman is a liar (,)?* (DCPSE:DL-D06)

The examples are of the same kind as examples 5.22-23, which were uttered not in LLC but in FCSE Broadcast Interviews & Discussions. In the context of these examples, I argued that the frequent use of communication verbs in Broadcast Interviews & Discussions goes back to such interpretative uses, which often have an explanatory character and/or refer back to a previous utterance. The HCFA result suggests that this use of communication verbs is actually most typical of the LLC data. Similar to what has been argued in the context of the MDCA, these uses stand out in LLC because they were already frequently used in mid-20th-century spoken English – a period which was followed by major increases of activity/event verbs and also of intermediate verbs (cf. Figure 5.6). At the same time, progressives formed with communication verbs are only slightly more frequent in Broadcast Interviews & Discussions today than in the LLC data. Thus, relative to the other verbal uses, they were especially frequent about 50 years ago.

Finally, there are three anti-types that shall briefly be addressed: (xix) *FCSE + Spontaneous Commentary + communication + present*, (xx) *ICE-GB + Spontaneous Commentary + communication + present*, (xxi) *LLC+ Spontaneous Commentary + communication + present*. Again, the analysis has identified three identical patterns that behave in the same way in the different subcorpora. In all three cases, the observed frequencies are way below the predicted levels of use.

A look at the actual tokens reveals that some of them are used interpretatively (similar to the interpretative uses in Broadcast Interviews & Discussions). They are illustrated in example 5.90-92:

(5.90) *Of course we**'re exaggerating** the effect here (,) so that it's easier for you to see („) („).* (DCPSE:DL-F09)

(5.91) *I think that**'s asking** a little bit much of him.* (DCPSE:DI-F13)

(5.92) *Well I think what they**'re saying** is, basically, that we're gonna attack better than you are.* (FCSE_SC_2_110)

However, a much larger proportion of tokens is used descriptively to report what is going on in a match, parade, etc. These uses are highlighted in the following examples:

(5.93) *Her Majesty**'s speaking** to him now (,) at the end of the (,) line of presentations on the admiralty side.* (DCPSE:DL-F07)

(5.94) *He**'s signalling** that there were two men inside the five yard area („).* (DCPSE:DI-F04)

(5.95) *They're, they**'re not shouting** and waving a flag because somebody's told them to.* (FCSE_SC_7_261)

The preponderance of uses like these shows that the combination of the present progressive with communication verbs is used differently in Spontaneous Commentary than in Broadcast Interviews & Discussions. While interpretative uses figure prominently in the latter genre, uses in the former genre are much more descriptive – i.e aspectual instead of subjective. This relative lack of interpretative uses is a potential explanation for the constant underrepresentation of the pattern in Spontaneous Commentary. Furthermore, also the descriptive uses are comparatively infrequent if the actual token counts are considered. Very likely, this is due to the fact that it is mostly non-communicative activities and events that are described by sports/ceremonial commentators (cf. the high share of activity/event verbs in Spontaneous Commentary, Figure 5.6).

5.3.2. Conclusion

Altogether, the HCFA has confirmed and further refined the results presented in chapter 5.1 and 5.2. We have seen that in three of four genres, highest levels of progressive use are indeed connected to an overrepresentation of the most prototypical of all progressive uses, present progressives formed with activity/event or process/aspectual verbs. Likewise, a decline of the progressive's frequency in FCSE Face-to-face Conversation goes along with less frequent than expected use of this prototypical pattern. Thus, the proposed connection between the progressive's overall frequency and the development of its most prototypical uses receives further support.

Furthermore, the analysis has shown that variable combinations involving intermediate verbs are overrepresented in FCSE whereas they are underrepresented in LLC and ICE-GB. This supports the claim that such – often non-obligatory – progressive uses have become more common in recent years. State verbs, on the other hand, do not show up among the types and anti-types, corroborating the finding that their use has remained relatively stable across the three subcorpora.

Importantly, the HCFA has revealed types and anti-types that correspond to specific functional uses of the progressive. Some of them are infrequent and – arguably – of little

importance, while others, such as the co-occurrence of past progressives with communication verbs in Face-to-face Conversation, are relatively frequent and point towards aspects of progressive use that deserve closer examination and that have been addressed before (cf. Leech's (2004) observation on the recentness progressive as a feature of Face-to-face Conversation).

From a usage-based, constructionist angle, the HCFA results highlight that a general-level construction such as the progressive comprises many different constructional subschemas, whose use is sensitive to time and genre. It has been shown which of these constructional realisations have developed in a significant way – leading to over- or underrepresentation. These changes affect the progressive's constructional network and – in addition to the findings of chapter 5.1 and 5.2 – help us describe and understand the construction's recent development in spoken English.

5.4. Discussion of results

At this point, I would like to adopt a slightly broader perspective and come back to questions that were raised in chapter 2 and 3: Is the progressive changing under its own momentum? How do conventional and innovative uses interact? Can frequency effects be observed? In how far do the observed changes qualify as constructional changes? How does the usage-based CxG account help to interpret the results? Is the progressive developing towards a general imperfective marker? And what role does colloquialization play in the observed developments?

Hypothesis 1 assumed that the progressive's frequency increase was still under way in the 21st century. The frequencies reported in Figure 5.1 and Table 5.1, as well as in Figure 5.2 and Table 5.2 support this assumption. The only genre for which the results are not entirely straightforward is Face-to-face Conversation. There, the DCPSE and FCSE data suggests decreasing progressive use between the 1990s and today (following a frequency increase between the 1950s and the 1990s). However, the genre must be more variable than is generally assumed[37], which is indicated by the results form BNC-DS and Spoken BNC2014, suggesting an ongoing frequency increase in conversation for the same period.

As has been shown, the different levels of progressive use in ICE-GB and FCSE Face-to-face Conversation cannot be explained by different shares of narrative passages across the subcorpora (cf. Figure A.3 & A.4). However, it has been demonstrated that the frequency difference is linked to a relative decline of the most prototypical pattern of progressive use, combinations of the present progressive with activity/event and process/aspectual verbs. These uses are less frequent/typical in FCSE Face-to-face Conversation than in the corresponding ICE-GB genre, where they are exceptionally frequent and responsible for the high level of progressive use.

In Broadcast Interviews & Discussions, Parliamentary Language & Prepared Speech and Spontaneous Commentary, the progressive exhibits a continuous and significant

[37] A discussion of this claim is provided in chapter 7.

frequency increase. Compared to Face-to-face Conversation, all these genres have traditionally been more heavily influenced by various norms of written language.

In the context of Hypothesis 2, it was argued that the progressive's frequency increase is partly happening under its own momentum, as proposed by Leech et al. (2009: 269f.). Assuming that "repetition strengthens memory representations for linguistic forms and makes them more accessible" (Bybee 2007: 10), it was claimed that increased use of the progressive is linked to more frequent use in those contexts in which the construction has long been firmly established, resulting in further entrenchment of these prototypical uses. The corpus data largely confirms this prediction: Whenever a frequency increase can be observed (i.e. between LLC and ICE-GB Face-to-face Conversation and continuously from LLC to ICE-GB to FCSE in all other genres), the normalised frequencies of the present progressive and of progressives co-occurring with activity/event verbs significantly increase as well (cf. Figure 5.3 & 5.6). The reason why this increase per htw does in most cases not translate into a significantly higher level of use *relative* to the other morphosyntactic or verbal contexts is that progressive use in these other contexts (most importantly with modal auxiliaries and cognition and perception/sensation verbs) has increased as well. While this increase in contexts other than the most prototypical ones is less pronounced in terms of htw, it is sometimes more pronounced percentage-wise. Thus, also the second part of Hypothesis 2 is confirmed, which assumes that less typical, innovative uses increase as well (however, on a lower level). Even though progressive uses with intermediate verb classes, modal auxiliaries or *to*-infinitives do not constitute real linguistic innovations, they can be seen as innovations in a statistical sense: Contexts in which the progressive was never very frequent are slowly becoming more and more routinised.

Further support for a continuing expansion of prototypical uses in the context of a frequency increase was provided by the results of the HCFA: Except for Broadcast Interviews & Discussions, the variable combination present progressive plus activity/event or process/aspectual verb is overrepresented in all subcorpora and genres in which progressive frequencies are highest. For Face-to-face Conversation this is in the 1990s, for Parliamentary Language & Prepared Speech and Spontaneous Commentary it is today.

Regarding the development in Face-to-face Conversation, it seems unlikely that uses of the activity_process + present pattern had reached a saturation point towards the end of the 20th century – especially if the continued frequency increase of the progressive in the two BNC corpora is considered. Thus, it appears more likely that the ICE-GB conversational data simply contains an exceptionally high share of these prototypical progressive uses, which is not representative of the time period and which dwarfs present-day usage in these contexts.

Regarding the development of conventional and innovative uses, the corpus analyses can be summarised as follows: While in three of four genres the most prototypical uses have become still more frequent (per htw), an increased level of use with cognition and perception/sensation verbs can be observed in all genres. Also modal uses and combinations with *to*-infinitives show an upward trend. The same is true for the *BE being adj.* pattern. Furthermore, the MDCA has shown that progressive *have* is distinctive of present-day spoken English (in genres other than Face-to-face Conversation). Thus,

there is very frequent conventional use on the one hand, and less frequent but increasing use in less well-established/conventional contexts on the other hand. This is perfectly in line with De Smet's (2016) model of language change, which proposes an interaction of convention and innovation: "the more readily retrievable a conventional use of an expression is, the better are its chances of being used also in similar but unconventional ways. [...] This way, it appears, expressions' usage can continually expand to the fringes of what is grammatically conceivable" (2016: 86ff.). Using this model of change, it is possible to account for both aspects of the progressive's development – the supposed change under its own momentum and the increasing frequency of less typical, often optional and non-aspectual uses.

This interaction of conventional and innovative/less conventional uses brings us to the domain of frequency effects. In fact, it seems to represent a kind of attraction effect, whereby frequent use of the progressive leads to an "attraction" of less prototypical uses. In all genres, also in Face-to-face Conversation, the progressive's frequency is significantly higher today than in mid-20th-century English. In the light of De Smet's model of change and of what is known about frequency effects (cf. chapter 2.1.2), it seems perfectly justified to propose that more frequent use of, for example, progressive *expect*, *feel* or *hope* is a result of this generally increased cognitive availability of the progressive construction. Similarly, increasing progressive use of certain polysemous verbs, e.g. *have* or *look*, could be due to the same effect: As aspectual, dynamic uses of these verbs become more frequent (as a result of the general frequency increase of the progressive), also non-aspectual, optional uses become more likely, since they differ only along certain semantic lines. Increasing progressive frequency leads to a widening of certain aspects of the construction's semantic and morphosyntactic paradigm. Importantly, this diversification must not be confused with a replacement of the prototypical uses by the supposedly innovative ones. Rather, it should be seen as an addition that is affecting specific parts of the constructional network.

Another frequency effect is pragmatic strengthening (or pragmaticalization) of specific progressive uses. As has been argued in chapter 2.1.2, pragmatic strengthening is based on repetition but does not normally require very high frequencies since it is repetition in appropriate contexts that is most important. The HCFA results showed that past progressive use with communication verbs is typical of Face-to-face Conversation in all three subcorpora. As qualitative analysis of the respective tokens revealed, many of them qualify as recentness progressives. As Leech (2004: 32) has pointed out, the recentness progressive is a feature of conversational style that cannot easily be explained by recourse to the progressive's semantics. Instead, frequent use of the pattern in the appropriate contexts has resulted in the notion of recentness and the distinct conversational tone that accompanies its use.

Similarly, the HCFA type *LLC + Broadcast Interviews & Discussions + communication + present* was identified as a pattern that is often used with interpretative meaning in argumentative/explanatory contexts (example 5.87-89). Again, we observe a (partial) correspondence between a form that is frequently used in a given context and a pragmatic function it expresses.

Another example of a progressive form/pattern that has acquired a clear pragmatic

meaning is the *BE being adj.* pattern, which is used to refer to deliberate/specific behaviour. Also the so-called matter-of-course future is a case in point: frequent use of *will* plus progressive in specific semantic contexts has resulted in the pragmatic inference that the described future event will happen as a matter of course. This meaning is well-established in present-day usage and most likely further increasing in frequency (cf. chapter 5.2.1).

A crucial question in the context of usage-based CxG regards the nature of the observed change(s): Do they qualify as constructional change? First, they are construction specific – meeting the first of Hilpert's criteria (cf. chapter 2.2.2). Second, it has been shown that the progressive has developed at different rates in different genres, meaning that its distribution in the speech community has changed. According to Hilpert's definition (2013a: 16), such distributional changes constitute constructional change. Third, changes in the progressive's semantic, lexical and morphosyntactic paradigm have occurred. Irrespective of the question whether these changes constitute grammaticalization, colloquialization or something else, they clearly qualify as constructional changes since they reflect changed contextual "preferences" of the progressive. As has been pointed out before, it is the relative frequency changes that Hilpert regards as especially indicative of constructional change (2013a: 17). Thus, the fact that progressives with intermediate verbs are overrepresented in three of four genres (cf. Figure 5.7) can be interpreted as a change in the cloud of progressive exemplars in speakers' minds. Similarly, changes in the construction's distinctive collexemes also qualify as instances of constructional change. Finally, the HCFA has shown that specific configurations of progressive use are typical or atypical in different subcorpora and genres. If these configurations are regarded as constructional subschemas, the observed variation in their diachronic distribution corresponds to changes in the progressive's constructional network. For example, the subschema present progressive plus intermediate verb is atypical of Broadcast Interviews & Discussions in the 1990s but typical of progressive use in the same genre today. At the same time, other schemas remain stable over time; most notably the combination of past progressives with communication verbs in Face-to-face Conversation, which is frequently realised as the so-called recentness progressive.

In the light of this evidence, it can be concluded that the progressive's recent development in spoken British English does indeed constitute a case of constructional change. What is less clear, however, is whether (some of) these changes also qualify as grammaticalization, or – more specifically – whether they are indicative of the frequently proposed development of the progressive towards a general imperfective marker. As has been explained in chapter 2.2.2, many frequency changes that qualify as constructional changes are not part of grammaticalization processes (Hilpert 2013a: 12). I have also made reference to Mair (2004), who argues that the exact relationship of frequency changes and grammaticalization can be very complex and is in many cases far from clear. However, Mair also argues that it is often relative frequency changes (as opposed to changes per htw) that are indicative of ongoing grammaticalization processes. In the course of these changes, one variable is replacing another, transforming the profile of a given construction (2004: 138). While relative frequency changes of the progressive's verbal/semantic paradigm have been observed that could be indicative of a beginning

drift of the construction towards a general imperfective maker (relative increase in intermediate verb classes in three of four genres, overrepresentation of individual progressive verbs such as *have*, *think* or *feel*, intermediate verbs occurring among FCSE types and among LLC and ICE-GB anti-types), these changes are clearly not crowding out the prototypical aspectual uses. On the contrary, for three of four genres, increasing use of the present progressive and of progressives co-occurring with activity/event verbs has been reported. Most importantly, progressive use with stative verbs shows little diachronic variation. In Spontaneous Commentary, it has even developed from (significant) over- to (non-significant) underrepresentation (cf. Figure 5.7). If the progressive was developing into a continuous or imperfective aspect, one would clearly expect stative verbs to be on the rise at the expense of the established aspectual uses. Since this is not the case, a wholesale progressive-imperfective drift seems very unlikely. Consequently, the observed constructional changes do not appear to reflect ongoing grammaticalization of the progressive construction. Instead, the progressive's functional core has remained comparatively stable, while – at the same time – the construction's use has diversified and spread to certain non-prototypical contexts in which it is becoming more and more established.

Finally, one important sociolinguistic concept has to be addressed: colloquialization. As has been explained in chapter 3, it goes back to Mair (2006b: 183ff.) and proposes a stylistic shift from the written to the spoken norm in contemporary English. Smith (2005) thoroughly analysed the role of colloquialization in the progressive's development in written 20th-century English and concluded that it was one of several factors in the construction's frequency increase. The results of the present study show that the progressive's frequency increase has gained considerable momentum in spoken genres which have traditionally been influenced by the norms of written language. In two of these genres, Broadcast Interviews & Discussions and Spontaneous Commentary, today's levels of progressive use have even surpassed the level in Face-to-face Conversation. Since the progressive is generally regarded as a feature of spoken language, where it reaches higher frequencies than in the written medium, this spread from the most natural of spoken genres to more regulated/formal discursive contexts can be regarded as an instance of ongoing colloquialization. Even though the progressive is not a marked or very salient linguistic feature, its increasing use in, for example, formal interview contexts – possibly even with subjective, interpretative function – can nonetheless be taken as an indication that stylistic norms have changed in recent decades. Thus, in addition to language internal explanations such as the construction's supposed change under its own momentum and the assumed connection between conventional and innovative uses, colloquialization constitutes a further – socio-cultural – explanatory factor in the analysis of the progressive's frequency increase.

6. Frequent patterns of use: Analysing pragmatic meaning and routinisation

After analysing "classic" variables of progressive use in chapter 5 – the frequency development in different genres, as well as the impact of structural and semantic factors – the present chapter moves on to focus on an aspect of variation that has received much less attention in previous research: (semi-)fixed patterns of use and their recent development. This means that from a constructionist perspective, the analytical focus is put on lower level schemas of the progressive that keep re-occurring in the data with considerable frequency and that – as will be argued in the course of this chapter – often have become discursive routines with specific discourse-pragmatic functions.

6.1. Lexical-grammatical patterns: State of the analysis

The progressive expresses a diverse range of functions, which – as argued in chapter 3.4.4 – are in many cases connected to specific lexical-grammatical patterns. A well-known example is the *matter-of-course future*, where the combination *will/shall* + progressive is used to refer to future events in their entirety (i.e. as perfective) and as happening independently of external influence (example 3.23) (Leech 2004: 66ff.). Other examples include the *ALWAYS*-type progressive, expressing a subjective evaluation of a situation (example 3.14) (Kranich 2013: 17), and the *BE being adj.* pattern, which makes reference to wilful temporary behaviour (example 3.8) (König 1995a: 157).

In addition to that, the analysis in chapter 5.3 has brought into focus complex genre- and period-specific uses (i.e. variable combinations) of the progressive that – in some cases – correspond to specific functions. For instance, it has been shown that progressives combining communication verbs and the past tense are significantly overrepresented in Face-to-face Conversation in LLC, ICE-GB and FCSE. Many of these tokens qualify as recentness progressives, making reference to a recent and completed communicative event and conveying an informal tone (cf. examples 5.77-5.82) (Pfaff et al. 2013). If such variable combinations are regarded as patterns of use, their correspondence to special discourse-pragmatic functions can be taken as further evidence that many of these functions come with genre-specific, morphosyntactic, semantic and/or lexical restrictions.

A central study in this respect is Römer (2005), who argues that the progressive should not be regarded as a purely grammatical but as a lexical-grammatical phenomenon, which comprises many semi- and fully fixed subschemas/patterns. She stresses that many of these lexically specific patterns are closely linked to the progressive's complex functionality. For example, she identifies the pattern *I BE (just) wondering*, which occurs several dozen times in her corpus data (2005: 126,150). According to her, "[t]he

shade of meaning expressed here is clearly politeness or softening" (2005: 126) (i.e. what is called pragmatic uses in chapter 3.4.2). While the softening effect is enhanced by the adverb *just*, the personal pronoun *I* reflects the fact that we are dealing with a subjective use of the progressive. It is exactly along such lines that Römer works out the lexcial-grammatical nature of the construction.

However, Römer's work is strictly synchronic and the majority of patterns she describes are not thoroughly analysed from a quantitative perspective. Furthermore, while her work obviously qualifies as usage-based, it does not explicitly address most of the concepts introduced in chapter 2 of the present work. The current chapter seeks to do exactly this. It identifies frequently occurring progressive patterns and analyses their recent diachronic development from a quantitative and qualitative perspective, focusing on pragmatic meaning and degree of routinisation. The analysis is conducted in the conceptual framework of usage-based CxG, taking into account possible frequency effects. In doing so, it tries to further illuminate the progressive's state of development in spoken British English and aims to connect with the results of previous studies and the ones presented in chapter 5.

In chapter 3.5.2, I formulated the hypothesis for the analysis of lexically specific progressive patterns. For convenience's sake, it shall be repeated here:

Hypothesis 3 The progressive has to be analysed at different levels of complexity and abstraction (structurally as well as functionally). Frequent use of lexically specific progressive patterns in specific contexts is linked to the development of distinct pragmatic meanings and to increasing routinisation.

The hypothesis acknowledges the role of context in the process of pragmatic strengthening and routinisation: It is only in specific discursive environments that repeated use may finally result in these developments (cf. Bybee 2006: 721 & 2015: 133ff.).

6.2. Some theoretical considerations

The study of recurrent patterns of use that are partially or fully lexically specific has attracted considerable attention in recent years. Related phenomena have been studied under headings such as *collocation*, *formulaic language*, *idiomatic expressions*, *routines*, *multiword phrases/units*, *lexical patterns/clusters/bundles/prefabs, sentence stems*, etc. A seminal publication is Sinclair's *Corpus, Concordance, Collocation* (1991), proposing the so-called *idiom principle*, which states that "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments" (1991: 110). This principle has inspired countless subsequent publications and is still a central topic of debate in many areas of linguistics (cf. Siyanova-Chanturia and Martinez 2015 for a recent review). Further influential publications include Pawley and Syder (1983), Nattinger and DeCarrico (1992), Aijmer (1996), Hunston and Francis (2000), Wray (2002) and Biber

et al. (1999, chapter 13).[1]

The study of formulaic language is of central interest to linguists working in the usage-based, constructionist paradigm. Since CxG does not posit a dichotomy between grammar and the lexicon, it seems perfectly suited to study patterns of language use that show aspects of lexical as well as grammatical items. As a theory of language that aims at cognitive plausibility, usage-based CxG is – among other things – concerned with the question of processing and mental representation of lexically specific patterns. This entails a focus on aspects such as frequency of use, schematization, chunking, entrenchment and routinisation, etc. As will be seen, however, also aspects such as discursive context and pragmatic associations play a central role in the understanding of recurrent patterns of use (cf. Schmid 2014).

Let us first focus on the issue of frequency and mental representation: Typically, it is in corpus data where recurrent patterns are identified. Aijmer (1996) and Biber et al. (1999, chapter 13), for example, identified an enormous range of discursive routines that speakers of English use in conversation (e.g. *how are you*, *thank you*, *I think*, *can you* ____, etc.). While it is beyond doubt that such patterns facilitate language production and comprehension, their actual mental status is often far from clear (cf. Blumenthal-Dramé 2017). Cognitively oriented corpus linguists have frequently argued that such patterns – if they are frequent enough and/or not fully compositional in meaning – must be stored holistically, i.e. as chunks. As such, they would have acquired construction status (cf. Bybee and Scheibman 1999; Bybee and Moder 2017; Kay and Fillmore 1999; Rödel 2014). However, based on recent results from psycholinguistic research, the matter appears to be less straight forward. What clearly seems to be supported is the claim that frequently occurring combinations are processed faster than infrequent ones (cf. Arnon and Snider 2010; Bannard and Matthews 2008; Caldwell-Harris et al. 2012; Kapatsinski and Radicke 2009; Tremblay et al. 2011). Regarding the question of holistic storage/chunking, the picture is more complicated: While Tremblay et al. (2011) found evidence for holistic storage of lexical bundles (e.g. *I mean if you* or *it was going to*), Schmitt et al. (2004) arrived at inconsistent results. Using an oral response task, they found considerable differences for the tested corpus-derived clusters and for the different participants. Consequently, they conclude that "corpus data on its own is a poor indicator of whether those clusters are actually stored in the mind as wholes" (2004: 147).

In the light of these findings it seems safe to conclude that frequent use of formulaic patterns is indeed indexed in speakers' mental grammars, which is in line with the concept of exemplar representation (cf. chapter 2.1.1). However, this does not necessarily entail that such patterns have to be processed holistically as chunks (cf. Schmid 2010). While this will most likely be the case for very frequent and conventionalised patterns, it will not apply to many other patterns that occur with middle or low corpus frequencies and that are less routinised in the language system. This view is in line with Schmid

[1] Among the more recent publications, the following should be mentioned: Biber et al. (2004); Christiansen and Arnon (2017); Corrigan et al. (2009a & 2009b); Culicover et al. (2017); Ellis and Ogden (2017); Herbst et al. (2011); Herbst et al. (2014); Römer and Schulze (2009); Schmitt (2004); Schneider (2014); Simpson-Vlach and Ellis (2010); Wood (2015); Wray (2008, 2012, 2017).

(2010), who also calls for cautious interpretation of corpus data regarding questions of mental representation and processing.

In addition to frequency, another crucial aspect of (semi-)fixed patterns concerns the pragmatic associations they can acquire. Schmid (2014) provides an insightful assessment of the pragmatic status of frequently occurring *lexico-grammatical patterns*. These are defined as "recurrent sequences of lexical and grammatical elements which serve an identifiable function" (2014: 254f.). The term comprises a number of different routines that can be fully fixed or relatively variable. There are, for example, lexical bundles, which are sequences of words that commonly co-occur in natural discourse, irrespective of their idiomaticity and internal structure (cf. Biber et al. 1999: 990ff.).[2] Aijmer (1996) studied them using the term conversational routines. They include phrases such as *I don't think so*, *you don't have to*, or *what do you think* (Biber et al. 1999: 990ff.). Many lexical bundles have acquired distinct pragmatic meanings. For example, the phrase *I don't think so* is commonly used to express disagreement, while *you don't have to* can be interpreted as an act of giving permission, and *what do you think* expresses an inquiry about the hearer's point of view (Schmid 2014: 271f.). It is at this point that we can observe similarities with many of the lexical-grammatical progressive patterns that have been identified by Römer (2005). *I'm just wondering*, for example, is frequently used as a conversational hedge and, thereby, qualifies as a lexical bundle in the sense of Schmid (2014). According to him, "pragmatic associations are likely to be instrumental, or even play a central role, in the acquisition of syntagmatic chunks" (2014: 254). Pragmatic meaning and frequency/routinisation seem to be connected reciprocally: On the one hand, lexical-grammatical patterns that have acquired certain discourse-pragmatic associations are more easily learned and more likely to become routinised in the speech community. On the other hand, frequent use in specific contexts contributes to the development of specific discourse-pragmatic functions (Bybee 2006: 721; 2015: 133ff.; Kay and Fillmore 1999).

There are three further important aspects that need to be taken into account if one aims to understand the emergence and development of (semi-)fixed patterns of use: priming, alignment and routinisation. As explained in chapter 2.1.1, Pickering and Garrod's (2004) model of interactive alignment assumes speakers to align their use of linguistic expressions in dialogue. This alignment is based on priming and can result in transient short-term and stable long-term routines (Pickering and Garrod 2017). While the short-term effects of priming can be observed as recency effects (i.e. persistence) in natural conversation (cf. Szmrecsanyi 2006), stable routines can result in lexical bundles. By invoking the concepts of priming and alignment, it is possible to bridge the gap between the micro-level (dialogic situation) and the macro-level (community grammar) of language use, thereby arriving at a comprehensive account of *routinisation* (i.e. the establishment of recurrent discursive routines). This is why the present chapter will not only look at the frequency development and the pragmatic associations of patterns of

[2] Biber et al. (1999: 992f.) consider combinations up to four words as lexical bundles if they occur with a minimum frequency of ten times per million words. For five- and six-word bundles, they set the threshold at five occurrences per million words.

progressive use, but will also take into account possible priming and recency effects (cf. chapter 4.4.5 for the relevant methodological considerations).

Finally, I would briefly like to ask what role lexically specific progressive patterns can play in the construction's overall development as reported in chapter 5. Or, to put it like Bybee and Torres Cacoullos (2009): How do the particular and the general interact in language change? Consider their following statement:

> One of our interests [...] is to examine how specific exemplars of constructions affect the overall meaning and use of the construction. We cast this question in a diachronic context and examine the way [...] prefabs interact with the more general construction [...]. Rather than viewing prefabs as something distinct from and perhaps peripheral to grammar in the traditional sense, we argue that prefabs constitute important loci of grammatical development in the diachronic domain. By implication, such conventionalized expressions have important interactions with more general constructions in the synchronic domain. (2009: 190)

Acknowledging that "prefabs develop their own discourse-pragmatic characteristics" (2009: 192), the authors also argue that "units of formulaic language maintain associations with productive constructions, contra the view that would isolate the former in a lexicon separate from the grammar" (2009: 193).

All this shows that a frequent discourse-pragmatic progressive pattern may act as a locus for the verbal extension of the construction. For example, the progressive's increasing use with cognition verbs (cf. chapter 5) might spread from patterns such as *I BE just wondering*, first to very similar and later to less similar contexts of use. This way, increasing use of specific lexical-grammatical patterns contributes to the overall frequency development of the progressive in a noticeable way.

In the course of this chapter, I will analyse the frequency profile of selected progressive patterns and their different realisations. This will include a focus on normalised as well as relative frequencies and also on the progressive/simple aspect ratios of selected patterns. Such a quantitative approach will illuminate the patterns' status and development in the language system (normalised frequencies) and yield insights regarding the constructional status (relative frequencies). Furthermore, it will show how fixed the aspectual choice (i.e. the progressive encoding) actually is, including its development in recent decades (ratio progressive/simple). The analysis will also take into account possible priming and recency effects – i.e. focus on the question how previous aspectual encoding influences pattern choice.

In addition to this quantitative approach, the pragmatic aspects of the patterns will be studied qualitatively. Going back to the actual concordances, I will focus on salient pragmatic uses and analyse whether they can be attested consistently. If this is case, the analysis will move on to the question of compositionality, which will be taken as an indicator (but not as proof; see discussion above) of unit status. Finally, the issue of productivity will be addressed: Are there comparable verbal contexts in which a pattern is used with similar pragmatic meaning? If yes, how many? And how have they

developed in the course of the past twenty years? Taken together, the different analyses will allow an assessment of the patterns' degree of routinisation.

6.3. Corpus evidence

The present chapter will not only make use of DCPSE and FCSE but also of BNC-DS (the demographically sampled, i.e. conversational part of the BNC's spoken section) and the recently released Spoken BNC2014 (Love et al. 2017) (cf. chapter 4.2.3).[3] This way, it is not only possible to identify recurrent progressive patterns, but also to study their use and recent development in detail from a qualitative as well as quantitative perspective.

6.3.1. Identifying patterns

At the beginning of the study of (semi-)fixed patterns of use stands the question of identification. How can recurrent patterns be identified in corpus data? Depending on method of identification, one will have to analyse different kinds of data. N-grams will look slightly differently from patterns (i.e. clusters of a certain size), word patterns will be very different from part-of-speech patterns, and even if one decides to analyse word patterns, the analyst still has to decide whether to sort them to the left or to the right. Since all these decisions result in different kinds of output, they influence the resulting analysis considerably. Or put somewhat simply: By adopting a particular method, one will get a particular kind of results, which will draw attention to patterns of a certain kind but divert attention from patterns of a different kind. While this is not problematic in itself, it should be kept in mind during the analysis and when comparing one's results to those of other studies.

After probing several ways of pattern identification, I decided to focus on four-word patterns (with the *-ing* participle in rightmost position) and to base the selection of the most frequent of these patterns on the complete progressive output from DCPSE and FCSE. Thus, out of 8,085 progressive tokens, the most frequent four-word patterns were extracted. In a further step, the most frequent semi-fixed progressive patterns were derived and the top-ranked ones selected for detailed study in BNC94 and BNC14 (see also chapter 4.3.3). Proceeding this way, it was possible to base pattern identification on a large set of progressive tokens and to undertake further analysis of the selected patterns in two corpora large enough to provide meaningful insights for comparatively specific phenomena. Since the patterns' frequencies are relatively low compared to the progressive's overall frequency and since the BNC94 and BNC14 data base is much larger than the one of DCPSE and FCSE, it was decided to report frequencies per million words (pmw) and not per hundred thousand words (htw) as in chapter 5.

The following patterns[4] are the result of this identification and selection process and will serve as the object of study in the present chapter:

[3] For convenience's sake, Spoken BNC2014 will from now on be referred to as *BNC14*. Due to its publication date, BNC-DS will be called *BNC94* (cf. BNC reference guide (Burnard 2000)).

[4] PP = Personal pronoun.

Pattern I I am/'m not *-ing*

Pattern II I BE just *-ing*

Pattern III what PP BE *-ing*

Pattern IV what BE you *-ing*

The decision to focus on four-word patterns was mainly based on the following reasons: First, in her study of lexical-grammatical progressive patterns, Römer (2005) also focuses on recurrent word patterns and not on part-of-speech patterns, N-grams, or some related phenomenon, which proved to be very useful. Other studies of conversational routines in English did so in a similar way (e.g. Biber et al. 1999: 987ff.). While I also retrieved three-word and five-word patterns, I decided to focus on four-word patterns because they turned out to be more interesting from a discourse-pragmatic perspective than most of the more frequent three-word patterns. Five-word patterns, on the other hand, would also have been a very interesting object of study. They, however, were too infrequent in the available corpora. Thus, a grain size of four words seemed most appropriate. That this decision is not unreasonable is indicated by publications such as Arnon and Snider (2010) and Bannard and Matthews (2008), which also studied four-word combinations, providing evidence for their cognitive reality.

6.3.2. *I am/'m not -ing*

The first progressive pattern to be studied in this chapter is the *I am/'m not -ing* pattern (cf. example 6.1-2). In BNC14, it occurs 1,751 times, reaching a frequency of 153 tokens pmw. In BNC94, it occurs 874 times, which amounts to a frequency of 174 tokens pmw. Thus, we are dealing with a pattern of considerable frequency, which has become slightly less common in recent decades[5] – counter to the progressive's overall development in the two corpora. While these numbers offer a first orientation, we need to analyse the pattern's frequency profile in more detail to be able to make more informed statements about its use and recent development.

(6.1) *yeah I know* ***I'm not saying I'm not saying*** *Christianity is any less weird than Islam.* (BNC14_S8Q3_170)

(6.2) *I'm not bothered* ***I'm not doing*** *anything Christmassy I haven't even got an advent calendar.* (BNC14_S4W8_31)

Frequency profile

Figure 6.1 depicts the normalised (pmw) frequencies of the pattern's 20 most frequent realisations in BNC94 and BNC14.[6]

[5] The decrease is significant for $p<0.001$.

[6] In descending order of the BNC94 frequencies.

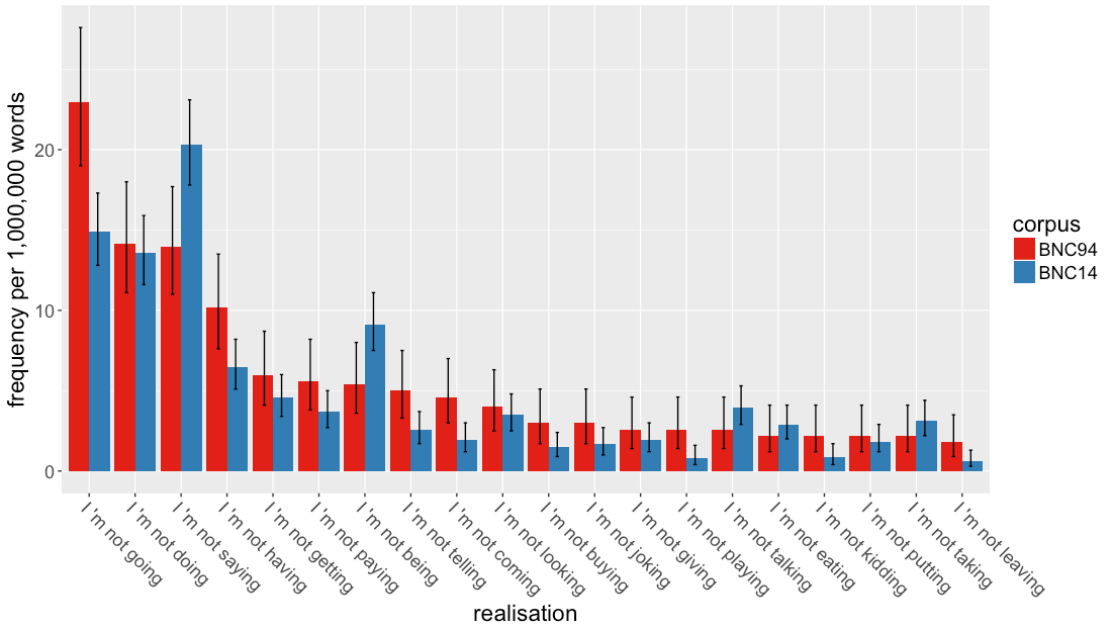

Figure 6.1.: Frequency profile of the *I am/'m not -ing* pattern (20 most freq. realisations with 95% confidence intervals) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

Three aspects of the plot are particularly salient: First, there are several realisations that are comparatively frequent and reach a level of 10 tokens pmw or higher. Second, the pattern's frequency profile has changed considerably in the course of the past twenty years. Most striking is the decrease of *I'm not going*[7], which was the pattern's most frequent realisation in Face-to-face Conversation of the early 1990s. Today, however, *I'm not saying* is by far the most frequent use – almost as frequent as *I'm not going* twenty years ago.[8] Third, all of the 20 most frequent realisations are formed with the contracted form *'m*. Given that we are dealing with conversational data, this does not come as a surprise. Altogether, only very few full forms of *am* occur in the data set.

As reported above, the pattern is less frequent today than it was twenty years ago. This downward trend is not just reflected in the massive decline of *I'm not going*, but also in the development of many other realisations: Of the 20 most frequent verb types, 15 have decreased in frequency. In five cases (*go, have, tell, come, leave*), this decrease is statistically significant. In comparison, only with five verbs does the pattern show an upward development, which is significant in only two cases (*say* and *be*). Interestingly, among the five significantly decreasing verbal realisations, there are two frequent activity/event verbs (*go* and *come*). Counter to that, no activity/event verb is among those two types with which the pattern has become significantly more frequent. Thus,

[7] *going to*-futures were excluded from the counts.

[8] Both changes are statistically significant. The exact significance levels for all realisations are listed in Table A.5 in the appendix.

the pattern's overall decline manifests itself on many levels, including typical progressive verbs. Significantly increased usage, on the other hand, is restricted to two verbs. In the face of a general downward trend, their pronounced increase might be indicative of a discourse-functional change/specialisation of the pattern. I will come back to this claim in the following section.

Since Figure 6.1 reports normalised frequencies, the results can be taken as an indication of the pattern's status in the language system, i.e. the community grammar. The development looks quite similar if relative frequencies – i.e. the frequencies of the different verbal realisations compared against each other – are considered (cf. Figure A.9). While minor developments are reversed (e.g. *do*, which exhibits a non-significant relative increase), the overall picture is very much in line with the one of the normalised frequency development. Perhaps the most noteworthy difference concerns the status of *say* in BNC14. If normalised pmw, its frequency is still lower than the one of *go* in BNC94. In relative terms, however, the two realisations are on exactly the same level. Obviously, this is an effect of the pattern's changed overall frequency (an absolute increase of *say* in the face of an overall decline boosts *say*'s relative frequency).

Specific realisations

The main focus of this section will be on the pattern's most frequent realisation in BNC14, *I'm not saying*. Additionally, I will take a look at *I'm not being*, which is the only other realisation that has become significantly more frequent over the past two decades. Finally, I will briefly consider *I'm not having*. Given that *have* is clearly no typical progressive verb, its comparatively high frequency in this specific context – which has, however, decreased in recent years – is somewhat surprising and deserves to be addressed.

I'm not saying The pattern's use with *say* is exemplified in example 6.1. Its larger context is provided below: Mother (S0417) and son (S0416) are involved in a discussion about religion, in which the son feels compelled to explain and defend his line of argument:

> *S0417*: (.) I'm alright you don't have to get defensive all I'm asking is (.) what do you know about different religions do you think? do you know anything about Islam?
>
> *S0416*: not really I know a bit it's a bit crazy Islam as well
>
> *S0418*: –UNCLEARWORD
>
> *S0416*: apparently their god
>
> *S0418*: yum
>
> *S0416*: their god Allah the one and only god
>
> *S0418*: » yum yum –UNCLEARWORD
>
> *S0417*: mm

> *S0416*: Allah the one and only god which can't be depicted told erm told Muhammed to go er go and kill his child on top of a mountain
>
> *S0417*: yeah well so did Christianity
>
> *S0416*: mm?
>
> *S0417*: it's the same thing Muhammed –UNCLEARWORD the same
>
> *S0416*: yeah I know **I'm not saying I'm not saying** Christianity is any less weird than Islam
>
> *S0417*: yeah
>
> *S0416*: I'm just saying they're all weird in my eyes (BNC14_S8Q3)

Progressive *say* expresses two aspects of meaning. First, it makes reference to an ongoing verbal exchange – conveying progressivity, incompletion and limited duration (i.e. the construction's core aspectual functions (cf. chapter 3.4)). This, however, does not fully capture the functional force of the example. In addition to its aspectual meaning, *I'm not saying* refers back to what has just been said, serving as an interpretation and justification of a previous statement. In the present case, the speaker tries to explain that what he just said about Islam – e.g. "it's a bit crazy Islam as well" or "Allah [...] go and kill his child on top of a mountain" – was not meant to say that Christianity was any better. This intention is further highlighted by the subsequent use of *I'm just saying*, which serves a similar purpose (cf. chapter 6.3.3). Thus, in addition to the aspectual meaning, use of the progressive also conveys an interpretative function that is employed to refer back to, explain, and justify a recently made statement. It is exactly in these contexts that *I'm not saying* is most common. Consider the following examples:

(6.3) *no no* ***I'm not saying*** *that (.) I'm just saying (...) I'm just saying like [...].* (BNC14_S6KV_141)

(6.4) *yeah* ***I'm not saying*** *it's easy you'd need to train people in it.* (BNC14_SJ88_324)

(6.5) *[...]* ***I'm not saying*** *that's completely false but it's definitely not completely true.* (BNC14_S784_472)

(6.6) *it not justified to fight against the Nazis of course it was* ***I'm not saying*** *that's analogous I'm just saying [...].* (BNC14_SZP6_515)

(6.7) *[...]* ***I'm not saying*** *I'm right and he's wrong but what I've said [...].* (BNC14_SNNG_860)

In all these cases, the same interpretation that was provided above can be applied. This shows that *I'm not saying* has acquired a stable discourse-pragmatic function on top of its aspectual reading. Such an insight is closely in line with Schmid's model of lexical-grammatical patterns and their pragmatic associations (2014: 273). Furthermore, it is in line with Hypothesis 3, which proposes a connection between pragmatic meaning and frequency in specific contexts of use: *I'm not saying* is a feature of spoken conversation

with a distinct pragmatic meaning and has become considerably more frequent in the course of the past twenty years. While it is hard to say whether frequency acts as a cause of pragmatic strengthening, or whether pragmatic meaning leads to more and more frequent selection, it is obvious that the two phenomena are connected – most likely reciprocally.

I'm not saying's development might also help to explain the decline of *I'm not going*. Unlike *say*, *go* is not normally used interpretatively in argumentative contexts. This is exemplified by the following uses:

(6.8) *well I could turn around to my parents and say piss off* ***I'm not going*** *and what do we do then?* (BNC14_S4TV_4206)

(6.9) ***I'm not going*** *out to theatre openings every night.* (BNC14_S6W8_1819)

While *I'm not going* in example 6.8 qualifies as futurate progressive (cf. chapter 3.4.3)[9], the token in example 6.9 refers to an habitual action (cf. chapter 3.4.1). In neither of the two cases does it fulfil a discourse-pragmatic function similar to the one of *I'm not saying*. Consequently, if one assumes that due to the frequency increase with *say*, the *I am/'m not -ing* pattern has become more and more associated with interpretative contexts, it seems plausible that other uses are losing ground. However, if this was the only explanation, one would expect a similar decline for the use with *do*, which is more or less stable, though. Thus, the rapid decline of *go* might well be conditioned by a further factor: increasing restriction of *be going* to contexts of the *going to*-future. Assuming that the latter is further gaining ground in informal spoken English, uses of *be going* outside this context might be becoming less common.

Coming back to *I'm not saying*, I would now like to move on to the question of compositionality. Can its pragmatic, i.e. non-aspectual, meaning be completely derived from its component parts, or does it go beyond the individual constituents? As has been explained, the interpretative function of the pattern can be traced back to the aspectual meaning of the progressive. It only works in the context of an ongoing conversational exchange, where one interprets what has just been said. This aspect of the pattern's use is transparent. However, the explicit interpretative act performed by the pattern does not fully follow form the use of the progressive. Since it refers back to a previous statement, one would conventionally expect use of the simple past (*I did not say*). In this case, though, the sense of present relevance and subjective involvement would be missed.

A further aspect of the pattern's status as conversational routine concerns aspectual choice: How frequent is progressive *I'm not saying* relative to its grammatical competitor *I don't say*? Can recent changes be observed? Even though this question disregards functional considerations, it can be taken as an indicator of *how progressivised* the pattern actually is. Is its use so much connected to the progressive that the simple aspect encoding is becoming marginalised?

[9] Nesselhauf and Römer (2007: 324) identified *I'm not -ing* as a frequent context for futurate progressive use.

In order to find out, a binary logistic regression model was fitted. It predicts the choice between progressive and simple aspect for the four most frequent verbal realisations of the *I am/'m not -ing* pattern, i.e. *say, go, do* and *have.*[10] Doing so, it was possible to compare the results for *say* to those of other frequent verbal uses. While aspectual encoding was the dependent variable, the four different verbs and the two corpora (BNC94&14) served as predictor variables. Additionally, predictor variables were included that capture the phenomenon of conversational persistence (cf. Szmrecsanyi 2005 & 2006), i.e. the fact that speakers' linguistic choices are sensitive to priming and recency. As explained in chapter 4.4.5, priming was operationalised as aspectual encoding of the pattern's previous realisation, while recency was entered in the model as the (logarithmic) distance in tokens to its previous realisation.[11] Operationalisation of priming was further refined by the distinction between production-to-production and comprehension-to-production priming. In line with earlier corpus priming studies (cf. Szmrecsanyi 2005 & 2006; Gries 2005, 2007 & 2011), previous use of *I'm not -ing* was expected to positively influence speakers' readiness to maintain progressive encoding in subsequent use of the pattern instead of opting for the use of *I don't V*. This priming effect was expected to fade away with increasing textual distance.

Based on a stepwise selection process, the model reported in Table 6.1 was obtained.[12] It reveals that aspectual encoding is significantly influenced by the choice of verb and corpus. These two variables are not independent of each other but produce a significant interaction. Furthermore, aspectual choice is significantly primed by aspectual encoding of the previous token ("prev_realisation"), showing a significant interaction with distance ("log_distance"), which means that the priming effect is indeed subject to a recency effect. While the difference between production-to-production and comprehension-to-production priming occurred as predicted (the former being stronger than the latter), it did not turn out as significant and was excluded from the model.

Regarding the choice between *I'm not saying* and *I don't say*, the model reveals a strong and increasing preference for progressive encoding (as depicted in Figure A.10, which reports the progressive probability for the different verbs in the two corpora). While ca. 80% of the pattern's realisations with *say* were encoded as progressives in BNC94, this value lies above 90% in BNC14. Thus, uses of *I don't say* are indeed very infrequent in spoken conversation compared to *I'm not saying.* This preference for the progressive has become significantly stronger over the past twenty years (as can be seen in the "corpus and verb" section of the model summary in Table 6.1, where the corresponding change in log odds and the odds ratios are reported). The second strongest progressive preference in our pattern is exhibited by *go*, followed by *do* and

[10] All simple aspect encodings were retrieved from BNC-DS and Spoken BNC2014 using regular expressions and/or POS-tags.

[11] Since these effects can only be at work within a conversation, i.e. within a corpus text (or textual division in BNC-DS), textual boundaries had to be taken into account in the modelling process. Furthermore, previous aspectual encoding and its textual distance could not be coded for the first realisation of the pattern in a given text. Consequently, these tokens had to be removed from the analysis, resulting in a considerably diminished data base.

[12] Implemented as glm(realisation ~ corp*lemma + prev_realisation*log_distance).

	Dependent variable: realisation (simple vs. prog)			
	B (SE)	lower	*Odds ratio* incl. 95% CI	upper
Intercept	−1.07 (0.28) ***	0.20	0.34	0.59
Corpus and verb (pairwise comparisons):				
BNC94(do) vs. BNC14(do)	−0.44 (0.23)	0.41	0.65	1.02
BNC94(go) vs. BNC14(go)	−0.42 (0.25)	0.41	0.66	1.06
BNC94(have) vs. BNC14(have)	−1.71 (0.23) ***	0.12	0.18	0.20
BNC94(say) vs. BNC14(say)	1.21 (0.39) **	1.56	3.34	7.15
Persistence:				
prev_realisation(prog)	4.04 (0.55) ***	19.82	56.55	170.83
prev_realisation(prog)*log_distance	−0.65 (0.11) ***	0.42	0.52	0.65
Model χ^2 = 1,067.69 ***				
Nagelkerke's R^2 = 0.52				
N = 2,212				

Table 6.1.: Binary logistic regression: Aspectual encoding of *I am/'m not -ing* (*go, do, say, have*) (*** p<0.001, ** p<0.01, * p<0.05).

have (which, in the BNC14 data, occurs with the progressive in only about 10% of cases). Other than *say, do, go* and *have* show a downward trend: their realisation in the *I'm not -ing* pattern has become less frequent compared to the *I don't V* pattern. However, only the change for *have* is statistically significant. Altogether, the regression results are in line with what has been reported above: Not only is *I'm not saying* the most frequent verbal realisation of the *I am/'m not -ing* pattern in BNC14, it is also the realisation that most strongly disfavours use in the corresponding simple aspect pattern and this dispreference has become significantly more pronounced over time. In contrast to that, other frequent verbal realisations show lower frequencies/a downward trend in all these respects, providing further support for the above-made claim that the *I am/'m not -ing* progressive pattern is becoming increasingly associated with *say*.

The regression model also reveals that aspectual choice is not just influenced by corpus and verb, but also by the aspectual encoding of the previous token and by the distance to it. As can be seen in Table 6.1, previous progressive use ("prev_realisation(prog)") of the pattern significantly boosts subsequent progressive choice. This boost can be interpreted as priming effect (cf. Szmrecsanyi 2006; Gries 2005). As textual distance increases ("prev_realisation(prog)*log_distance"), this effect diminishes significantly. These results show that speakers are sensitive to the pattern's aspectual encoding and align their linguistic choices accordingly (cf. Pickering and Garrod 2004 & chapter 2.1.1). What role this alignment might play in the routinisation of *I'm not saying* will be discussed below.

A final aspect in the analysis of *I'm not saying* concerns its productivity, i.e. the

question whether closely related verbal contexts can be found that serve the same, or at least a very similar discourse-pragmatic function. If the pattern, including its pragmatic associations, was not restricted to *say*, one could assume that it is – at least to a certain degree – productive and not just a fixed discursive routine but one showing signs of schematization. In a first step, the analytic focus was narrowed to the pattern's use with communication verbs only. In a second step, I focused on those uses that show clear similarities with *I'm not saying* and that occur at least a few times in the corpus data.

At least one pattern emerged that is consistently used like *I'm not saying*. This pattern is *I'm not suggesting* and can be found both in BNC94 and BNC14 – admittedly, though, with very low frequencies (2 tokens in BNC94; 7 tokens in BNC14). Its usage in spoken English has already been reported by Römer (2005: 166), who describes it as an emphatic/attitudinal use of the progressive. While this description is relatively general, it does not run counter to a description that regards it as a special kind of interpretative progressive, used to refer back to and qualify a previously made statement. Consider the following examples from BNC14:[13]

(6.10) *I'll have a bit of a (.) that's healthy [...]* ***I'm not suggesting*** *–ANONname's cooking is unhealthy.* (BNC14_SJH6_354)

(6.11) *nobody's gonna [...] know that quote [...] no I don't think you need to no* ***I'm not suggesting*** *you read it out but [...].* (BNC14_SP2Y_1093)

(6.12) *yeah (.) well of course I mean (.) I I 'm not (.) no (.) but* ***I'm not suggesting*** *that I've got a like a massive global standpoint on it all I'm saying is like in my opinion we should tax them [...].* (BNC14_SVCS_326)

These examples are very similar to example 6.1 and examples 6.3-7. All the semantic and pragmatic implications that have been shown to apply to the use of *I'm not saying* do also apply here.

In addition, also *I'm not talking* can be found in comparable environments. It is more frequent than *suggest* (13 tokens in BNC94; 45 tokens in BNC14) but does not seem to be restricted to one specific context of use. Nonetheless, examples such as the following are clearly similar to the ones reported for *say* and *suggest*.

(6.13) *S0196: [...] when you said class I thought you meant as in s- you know stature as in like you know they're here and I'm here S0192: well I no no no no* ***I'm not talking*** *about where I'm no I (.) why would I think that?* (BNC14_S5XD_1131)

We can therefore conclude that pragmatically motivated use of *I'm not saying* can productively be extended to verbal contexts that are semantically similar and that license the same pragmatic associations.

After this detailed analysis of *I'm not saying*, I would now like to address – somewhat more briefly – the pattern's uses with *be* and *have*.

[13] See also example 5.88 in chapter 5.3.

I'm not being Two aspects of *I'm not being*'s use are particularly noteworthy. First, its frequency is surprisingly high, reaching a level of almost ten tokens pmw in BNC14 (raw freq = 104), which is considerably above the level of use of verbs such as *come, look* and *talk*. Apart from *say*, *be* is the only verbal context of the *I am/'m not -ing* pattern that has become significantly more frequent in the course of the past twenty years. The second noteworthy aspect concerns its functionality. As has been explained in chapter 3.3.2, the *BE being adj.* pattern is used to refer to an action as momentary and deliberate behaviour (cf. König: 1995a). Levin (2013) has convincingly shown that this use of the progressive has become more frequent in recent decades. Similarly, the results provided in chapter 5.2.2 show that use of *BE being adj.* has become significantly more frequent between BNC94 (ca. 30 tokens pmw) and BNC14 (ca. 43 tokens pmw).

At first glance, *I'm not being* seems to be the negated version of the *BE being adj.* pattern. It is usually followed by an adjective and can occur in contexts as the following, where the adjective refers to some kind of impolite or undesirable behaviour:

(6.14) *S1: Don't be so fucking petty! S2:* ***I'm not being petty*** *Gav!* (BNC94_KC6)

(6.15) *S0687: mum don't be rude S0688:* ***I'm not being rude*** *but he's not well he's not got our sense of humour has he?* (BNC14_S4HW_648)

However, these are not (or no longer) the contexts in which *I'm not being* is most frequently used. Instead, by far the most common single adjective following the pattern is *funny*, which – unlike *petty* or *rude* – refers to a positive type of behaviour. While six out of 27 uses in BNC94 (ca. 22%) are formed with funny, it is 63 out of 104 uses in BNC14 (ca. 61%).[14] Furthermore, *I'm not being funny* seems to have acquired an idiomatic meaning. First, consider example 6.15, in which the pattern's use is similar to the uses in example 6.14-15.

(6.16) *S1: Now don't look funny now Willie! [...] S2:* ***I'm not being funny!*** (BNC94_KDN)

In the majority of cases, however, *I'm not being funny* is used with a different meaning:

(6.17) *S1: Janice thinks it's brilliant. S2: Yeah, but she's sort of really naive and really S3: Gets on my nerves sometimes. S2:* ***I'm not being funny****, but she's a white man's woman, don't you think so?* (BNC94_KPG)

(6.18) *S0084: and then we were talking about the riots in London and he was getting really agitated and he was saying his cousin (.) female (.) very petite S0083: mm S0084: was er is in the police works for the Met and was part of the riot erm the riot troops that came out S0083: mm S0084: and he was saying* ***I'm not being funny*** *right but women you know why do they put them in these positions?* (BNC14_S7NH_760)

[14] Its simple aspect counterpart *I'm not funny* does not occur in BNC94 and only two times in BNC14.

(6.19) *[...] my mum sat me down and she goes* ***I'm not being funny*** *–ANONname but all your housemates are gonna have a meeting without you and bitch about how messy you are.* (BNC14_SK8T_64)

(6.20) *you –ANONname* ***I'm not being funny*** *yes you are partly to blame you should have put a stop to it there and then.* (BNC14_SNNG_717)

In these contexts, the pattern has qualities of a focusing device and hedge. The speaker signals that he or she is addressing a potentially delicate topic and – at the same time – tries to avoid sounding impolite. The meaning could possibly be paraphrased by "I don't want to sound rude" or "I'm sorry to say this". This is in line with Baxter and Wallace's (2009: 421) description of the pattern as a "typically British idiom", which is used "to downplay the effect of a sensitive or non-politically correct comment". Not unlike *I'm not saying*, *I'm not being funny* is used as a meta-discursive device that refers back – or in this case points towards – a statement made by the speaker.[15] In this sense, a link to the progressive's interpretative meaning may seem conceivable. However, the pattern's meaning is much more opaque, i.e. idiomatic, than the meaning of *I'm not saying*. While example 6.16 refers to a momentary/deliberate behaviour and is in line with the generally attested meaning of the *BE being adj.* pattern, this is not the case for the uses in example 6.17-20. The meaning of *I'm not being funny* has gradually extended to the observed hedging contexts in which the sense of momentary behaviour is backgrounded or lost and in which *not being funny* means talking in earnestness.

Thus, the statistically significant increase of *I'm not being* between BNC94 and BNC14 is mainly driven by a highly pragmatic and idiomatic kind of use that differs from the generally acknowledged function of the *BE being adj.* pattern. Again, we can observe a lexically specific use of the progressive that is being increasingly associated with a special discursive context, in which it has developed a prominent pragmatic – and in this case very much non-compositional – meaning.

I'm not having Use of progressive *have* has already been addressed in chapter 5.2.2, where it has been identified as distinctive of present-day progressive use (FCSE) in genres other than Face-to-face Conversation. A diverse functional profile emerged, including dynamic, stative, deontic, and futurate uses. In the case of the *I'm not having* pattern, however, matters are somewhat different. First, its use decreased from 10.2 tokens pmw (raw freq = 51) in BNC94 to 6.5 tokens pmw (raw freq = 74) in BNC14. While this is no dramatic shift in absolute terms, it is nonetheless statistically significant. Second, *I'm not having* shows a clear functional preference for special future contexts. Speakers use it to express that they do not want something (to happen) in the near/imminent future. This implication can be more or less strong: While the uses in example 6.21-22 could be paraphrased as *I don't want*, the ones in example 6.23-25 convey a stronger sense of rejection of a future state of affairs, expressing complete unwillingness of acceptance.

[15] In American English, a related meaning – emphasising that what you are saying is true – is expressed by the progressive phrase *I'm not kidding* or its reduced form *no kidding*.

(6.21) *Right, Max if you can't shut up* ***I'm not having*** *you in here [...].* (BNC94_KNV)

(6.22) *we were walking past burnt out cars that weren't there the day before and they weren't even batting an eyelid and I was thinking* ***I'm not having*** *my kids thinking this is normal life [...].* (BNC14_SGAN_755)

(6.23) *S1: No* ***I'm not having*** *that. S2: What? S1:* ***I'm not having*** *that. I'll smack her me bleeding sen. S2: You won't touch her. S1: I chuffing will.* (BNC94_KCX)

(6.24) *well no (.) that's bollocks [...] no* ***I'm not having*** *that.* (BNC14_SU82_3991)

(6.25) *no I don't want to no I don't want anything to do with that [...]* ***I'm not having*** *it no* ***I'm not having*** *it no no no.* (BNC14_S2T6_2181)

Although the pattern's frequency has decreased, its meaning does not show any clear signs of change. In both corpora, the majority of uses is found in contexts similar to the ones described here (often followed by *it, that, this* or *any*). Even though these uses are less idiomatic than the *I'm not being funny* pattern, their meaning is not fully compositional. Used in the right contexts, they convey the pragmatic implication of (strong/emphatic) rejection, which does not fully follow from the analysis of the pattern's component parts. Unlike *I'm not saying* and *I'm not being funny*, *I'm not having* is no meta-communicative device pointing to a specific statement in the conversation. Instead, it can be seen as a special kind of futurate progressive, which has acquired the pragmatics of rejection.

Assessing routinisation

The analysis has highlighted several characteristics of *I am/'m not -ing*'s use and recent development. Regarding its realisations with *say, be* and *have*, it should have become clear that pragmatic associations are central to the pattern's discursive functionality. However, the investigation has not only focused on pragmatics but also on different aspects of routinisation. The pattern's overall frequency has declined in the course of the past twenty years and this decline is reflected in decreasing use of many verbs. In absolute terms, the most frequent realisation in BNC94, *go*, has lost most ground, but also other frequent verbal choices such as *have, tell* and *come* have significantly declined. From this global perspective, it can be concluded that the pattern is less strongly anchored, or entrenched, in the speech community today than it was twenty years ago.

However, it has also been shown that two verbal realisations, *say* and *be*, have become significantly more frequent – both in terms of normalised and relative frequencies. *I'm not saying*, which was the pattern's third most frequent realisation in BNC94, has become its most frequent realisation in BNC14. Apart from the frequency increase, its use is characterised by a distinct discourse-pragmatic meaning, which can be regarded as a special case of the interpretative progressive. The corresponding simple aspect encoding, *I don't say*, is very much dispreferred (even more so in BNC14 than in BNC94), indicating

that the pattern is indeed very much progressivised. By revealing the same discourse-pragmatic associations for *I'm not suggesting*, *I'm not saying* has been shown to be productive – at least to a certain degree. In the light of this evidence, one can assume a relatively high degree of routinisation.

This assessment is further supported by the fact that the reduced form *not saying* also occurs in the BNC data, arguably an effect of the pattern's high frequency (cf. Bybee 2015: 42f.,124f.). The following conversational excerpt serves as an example:

S0517: no (.) think I'm very competent thank you

S0571: **not saying** you're not competent

S0517: mm?

S0571: I was just trying to help you become calmer and

S0517: » no mm (.) I am calm now (BNC14_SN3D)

In one respect, the corpus data has provided a direct window in speakers' psycholinguistic reality: The logistic regression model has revealed priming and recency effects for the aspectual encoding of the pattern. If a certain textual distance is not exceeded, previous progressive encoding significantly boosts subsequent progressive encoding. This finding is in line with previous corpus priming research (Gries 2005; Szmrecsanyi 2005) and has been explained by Pickering and Garrod's (2004) model of interactive alignment, whereby speakers converge on their linguistic choices in dialogue.

Apart from *say*, the pattern's use with *be* also shows clear signs of increasing routinisation. First, *I'm not being* has become significantly more frequent. Second, it increasingly co-occurs with the adjective *funny*. While *I'm not being funny* is much less frequent than *I'm not saying*, its meaning is much more opaque, having reached a high degree of idiomaticity, which, in turn, can be read as a sign of chunking. Even though it is rare, the reduced form *not being funny* is attested in the corpus data as well, which – like reduced *not saying* – is a result of the pattern's increased use and routinisation.

The only two uses of *I am/'m not -ing* that have become more frequent in the past twenty years, *say* and *be*, derive their discourse-pragmatic meaning from their meta-communicative usage. In the face of the decline of uses with *go* and more or less stable usage with *do* or *get*, this development can be read as an increasing bias of the *I am/'m not -ing* pattern towards use in meta-communicative contexts. This interpretation is also in line with the trend observed for *have*. While *I'm not having* is normally used in contexts with a very prominent – and at least partly non-compositional – pragmatic meaning, this meaning is not related to meta-communicative progressive use. Instead, it should be seen as a special kind of futurate progressive.

We can conclude that even though the progressive's overall frequency has increased between BNC94 and BNC14 (cf. chapter 5.1), no such increase can be attested for one of its most frequent four-word patterns *I am/'m not -ing*. However, if the focus is narrowed to specific verbal realisations, the picture becomes more differentiated. For *say* and *be* a significant frequency increase has occurred that is accompanied by various phenomena indicative of routinisation. In the end, the pattern's overall frequency decrease might well be a result of its increasing restriction to meta-communicative contexts.

6.3.3. *I BE just -ing*

The second pattern to be studied is semi-fixed *I BE just -ing*, which can occur in the present as well as past tense. Use of its most frequent realisation in BNC14, *I'm just saying*, has already been exemplified in example 6.3 and 6.6. Other uses are found in the following two examples:

(6.26) ***I'm just trying*** *to work out what is really the best thing to do [...].* (BNC14_SJJK_149)

(6.27) *S0026: stop thinking you look cool with your gun S0179: no I'm not but* ***I was just thinking*** *it was really funny to see such a reaction.* (BNC14_SUKT_1378)

The pattern's development is unambiguous: While it occurs with a frequency of 95 tokens pmw in BNC94 (raw freq = 475), its level of use has risen to 142 tokens pmw (raw freq = 1,624) in BNC14 – an increase of ca. 50%.[16] Thus, unlike *I am/'m not -ing*, the development of *I BE just -ing* is in line with the progressive's overall positive trend between BNC94 and BNC14.

Frequency profile

Figure 6.2 depicts the development (pmw) of the pattern's twenty most frequent lexical-grammatical realisations.[17] In BNC94, the distribution is surprisingly flat. Not one realisation emerges that would stand out as particularly frequent. This picture changes dramatically in BNC14, where the frequency of *I'm just saying* has increased from 6.6 to 17.2 tokens pmw (+160.6%***) (cf. Table A.6, appendix). Similarly, *I'm just thinking* increases from 3.4 to 8.0 tokens pmw (+135.3%***) and *I was just thinking* from 7.0 to 12.1 tokens pmw (+72.9%**). Apart from these very pronounced changes, the pattern's paradigm has remained relatively stable; none of the other changes is statistically significant. This means that the observed overall frequency increase is mainly driven by the development of these three realisations, which, today, are much more strongly represented in the conversational grammar of English than twenty years ago.[18]

If we change the focus from normalised to relative frequencies, some notable differences can be observed (cf. Figure A.11). Most importantly, only the development of *I'm just saying* remains statistically significant. Even though both *I'm* and *I was just thinking* also increase relative to the other uses, their development fails to reach significance. If relative frequency is regarded as a proxy for typicality, only *I'm just saying* has become significantly more typical of the *I BE just -ing* pattern in the course of the past twenty years.

Two of the pattern's most frequent realisations in BNC94, *I'm just trying* and *going*[19], have remained almost stable pmw but have – noticeably but non-significantly – lost

[16] The change is significant for $p<0.001$.
[17] In descending order of the BNC94 frequencies.
[18] If the tense distinction is disregarded, the development proceeds somewhat more orderly – but no less dramatically for *say* and *think* (cf. Figure A.12 & A.13).
[19] All cases of the *going to*-future were excluded.

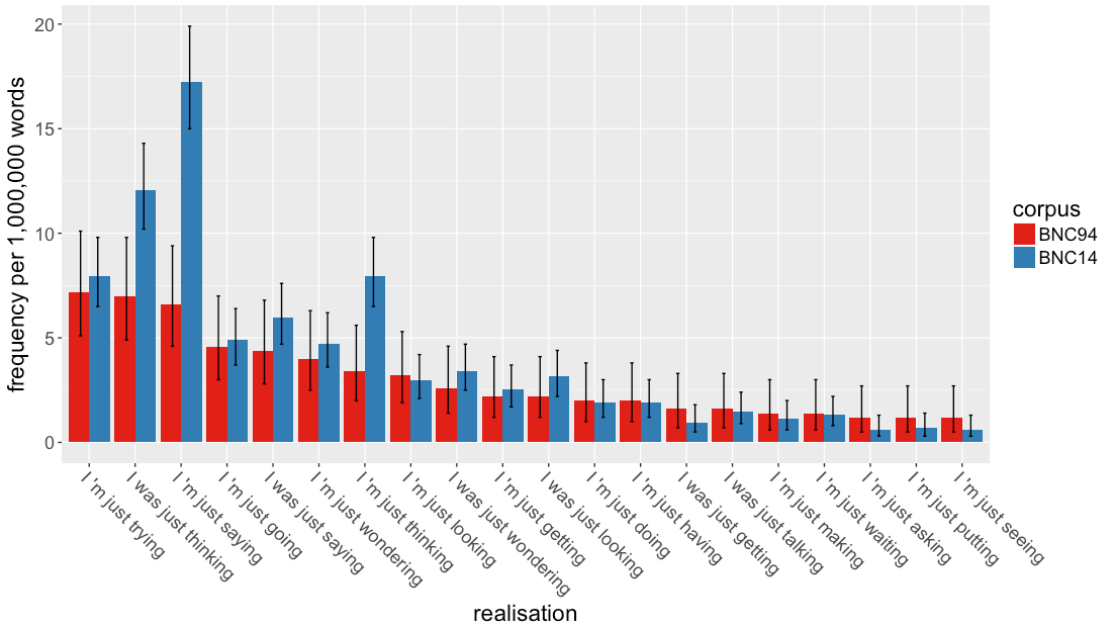

Figure 6.2.: Frequency profile of the *I BE just -ing* pattern (20 most freq. realisations with 95% confidence intervals) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

ground if measured relative to the other uses. Thus, while we see the pattern increasing with a communication (*say*) and a cognition verb (*think*), frequent uses with two activity/event verbs do show no such positive development and are – frequency-wise – left behind.[20]

Specific realisations

I will now focus on the pattern's use with *say* and *think* in more detail, addressing pragmatic associations and different aspects of routinisation.

I BE just saying If the tense distinction is disregarded, *say* already was the most frequent realisation of *I BE just -ing* in BNC94 – however, almost on the same level as *think* and only slightly more frequent than *try* (see Figure A.12 for an overview). Over time, its frequency has more than doubled – mainly driven by the present tense realisation *I'm just saying*.

[20]It should be noted, though, that the past tense realisation *I was just trying* has (non-significantly) increased from 1.2 tokens pmw (raw freq = 6) to 2.9 tokens pmw (rwa freq = 33). With six tokens in BNC94, it is as frequent as present tense *ask, put* and *see*, i.e. the pattern's realisations occupying rank eighteen to twenty in Figure 6.2 and Table A.6. However, since realisations with the same frequency were sorted alphabetically, past tense *try* is not included in the plot/table.

As the adverb *just* indicates, *I'm just saying* fulfils a downtoning or relativising function, rendering a following statement less definitive (cf. Römer 2005: 126). In the discussion of *I'm not saying*, we have already come across three examples where *I'm just saying* is used almost immediately after the negated pattern (example 6.3 & 6.6 and the longer conversational excerpt). In these examples, it hedges the expression of a personal opinion after the speaker has explained what he or she does not mean to say. Consider also the following examples:

(6.28) *S0115: [...] I saw it but it's one of them musicals that's sort of massively um anachronistic and S0037: what d'you mean? S0115: you know it's it's anachronistic (.) is it? [...]* ***I'm just saying*** *you know it's a bit like it's it's based on something but it's very liberal with like uh the facts.* (BNC14_SVKF_293)

(6.29) *S0592: yeah but I (.) yeah I don't think I would work in Mcdonald's though I wouldn't apply to work at Mcdonald's S0593: sometimes you have to start right at the bottom and work your way up S0592: I know* ***I'm just saying*** *I I would apply sooner to go and work in a grocery store rather than at Mcdonald's.* (BNC14_S8CB_236)

(6.30) *and then she was like guys and then and then I went (.) yeah cos it's just like a thing isn't it? people think that girls shouldn't do science and she was like what are you saying? girls should do science don't say that and I was like woah I was like no no no I wasn't saying that* ***I'm just saying*** *that it's like a stereotype [...].* (BNC14_S5QR_85)

In all these examples, the speaker explains his/her point of view by offering an interpretation of what he/she has said before – very similar to the use of *I'm not saying*. Again, the progressive is used as part of a meta-communicative expression that is based on the construction's interpretative function. Consequently, *I'm just saying* is not simply a downtoner or politeness strategy, but also serves a clear argumentative/explanatory purpose, making it a very useful linguistic device in spoken conversation. Additionally, a clear aspectual meaning component can be discerned. Since the pattern refers to an ongoing verbal exchange, the notions of progressivity and incompletion are not at odds with the interpretative reading.

The past tense realisation of the pattern is considerably less frequent and does not show a significant increase. Nonetheless, it is among the most frequent realisations of *I BE just -ing*. What does its functionality look like? Does it have the same pragmatic meaning as the present tense realisation? Even though subtle meaning differences are hard to distinguish objectively, I argue that there are two contexts in which *I was just saying* is used. In the first one, its meaning is very similar to the one attested for the present tense pattern:

(6.31) *no no I'm not saying that* ***I was just saying*** *that if (.) she had looks like older when he's on top of her he might not be thin- it might not occur to him that oh (.) this is actually the reality she's actually this young [...].* (BNC14_SXR9_1214)

(6.32) *S0192: what was that all about? told her she was a bad parent or something? S0189: no I wasn't saying that* ***I was just saying*** *like (.) how do I go into it? erm S0192: well don't talk about it if you don't want to.* (BNC14_SXQU_466)

In these examples, the speaker could just as well have used the present tense to explain what he or she actually said and meant. Use of the present progressive makes intuitively more sense, since it highlights the present relevance of the statement for the ongoing conversation. Rendered in the past tense, the fact that one refers to a previously made statement is highlighted. Since this statement is clearly completed, i.e. seen in its entirety, progressive instead of simple encoding is hard to explain from the perspective of aspectual semantics. Here, in addition to the interpretative function, the fact that we are dealing with a *recent* communicative event seems to license use of the progressive (cf. chapter 3.4.2 & 5.3 for examples of the so-called *recentness progressive*).

This recentness function is even more prominent in the second context of use, where the argumentative and interpretative force is absent:

(6.33) *UNKFEMALE: and you how are you? S0598: good good yeah I've got er* ***I was just saying*** *to –ANONname I've got tonsillitis just in time for Christmas but yeah apart from that I'm good yeah.* (BNC14_S6Q6_1164)

(6.34) *[...]* ***I was just saying*** *to dad that when we moved you know in many ways it was easier if you're the one going to work [...].* (BNC14_S8PW_1922)

Here, *just* is not used as a stance but as a temporal adverb. *I was just saying*'s pragmatic function is no longer that of an argumentative device but one of signalling recency (and, perhaps, of conveying an informal tone). It has been shown in chapter 5.3 that such recentness uses of the progressive are typical of Face-to-face Conversation (cf. example 5.77-82). In line with this interpretation, Biber et al. (1999: 1120f.) report an example of *I was just saying* in the context of their discussion of past progressive use with reporting verbs. In addition to recency, they also argue for an evidential meaning, with "the speaker vouch[ing] for the authority of her own message" (ibid.).

After this analysis of the pattern's discourse-pragmatic functions, let us briefly focus on the question of compositionality. Similar to *I'm not saying*, the meaning of *I'm just saying* can be linked to the progressive's interpretative function. Speakers use both patterns to explain what they really mean by a previously made statement. Together with the aspectual meaning components of present relevance and ongoingness, a partially non-compositional meaning emerges: While the aspectual associations are transparent, the act of referring to and interpreting a previous statement does not fully follow from the pattern's structure and is only established by repeated use in the appropriate contexts.

If the same interpretative meaning is expressed using the past progressive (example 6.31-32), the aspectual components of ongoingness and present relevance are lost since the pattern makes reference to a completed past situation. One motivation for progressive use is obscured and the pattern's meaning becomes more opaque, i.e. increasingly non-compositional.

Finally, there are the clear recentness uses (example 6.33-34), where the progressive does neither convey an interpretative function nor a notion of ongoingness. Instead, it is used in combination with the temporal adverb *just* to refer to a recent but completed communicative act. While *just* certainly helps to establish the notion of recency, none of the pattern's components indicates that the progressive's imperfective aspectuality is basically changed to a perfective one.

The investigation will now adopt a quantitative focus and address the pattern's degree of aspectual fixation. Is it clearly biased towards the progressive, or does the corresponding simple aspect use *I just say/said* also frequently occur in conversational discourse? In the same way as for *I am/'m not -ing*, a logistic regression model was fitted to predict the aspectual encoding. In addition to *say*, the four next-most frequent verbal realisations *think, try, wonder* and *go* were included in the analysis. In order to keep the model manageable, the tense distinction was disregarded. Model selection was undertaken in the same way as above, resulting in the model reported in Table 6.2.[21][22]

	Dependent variable: realisation (simple vs. prog)			
	B (SE)	lower	*Odds ratio* incl. 95% CI	upper
Intercept	−1.21 (0.45) **	0.12	0.30	0.71
Corpus and verb (pairwise comparison):				
BNC94(go) vs. BNC14(go)	−0.65 (0.40)	0.24	0.52	1.17
BNC94(say) vs. BNC14(say)	0.49 (0.32)	0.87	1.63	3.04
BNC94(think) vs. BNC14(think)	−0.56 (0.27) *	0.33	0.57	0.98
BNC94(try) vs. BNC14(try)	−1.50 (1.05)	0.03	0.22	1.77
BNC94(wonder) vs. BNC14(wonder)	0.99 (0.02) *	1.18	2.68	6.12
Persistence:				
prev_realisation(prog)	2.42 (0.50) ***	4.23	11.26	30.57
prev_realisation(prog)*log_distance	−0.34 (0.07) ***	0.62	0.71	0.82
prev_realisation(prog)*prev_speaker(same)	0.69 (0.24) **	1.25	2.00	3.21
Model $\chi^2 = 511.49$ ***				
Nagelkerke's $R^2 = 0.29$				
$N = 2{,}184$				

Table 6.2.: Binary logistic regression: Aspectual encoding of *I BE just -ing* (*say, think, try, wonder, go*) (*** p<0.001, ** p<0.01, * p<0.05).

Again, we get a significant interaction of corpus and verb as well as of previous aspectual encoding ("prev_realisation") and its textual distance ("log_distance"). As above,

[21] Implemented as glm(realisation ~ corp*lemma + prev_realisation*log_distance + prev_realisation*prev_speaker).

[22] Note that the model reaches a relatively small – but still acceptable (cf. chapter 4.4.5) – pseudo R^2 value (0.29), meaning that its explanatory power is lower than the one of the previous model, depicted in Table 6.1.

the latter is interpreted as a priming effect, fading away with time. Other than in the previous model, a further effect turned out as significant: the interaction of previous aspectual encoding and previous speaker ("prev_realisation*prev_speaker"), with previous speaker being either *same* or *different*. This interaction amounts to the difference between production-to-production (previous token uttered by same speaker) and comprehension-to-production priming (previous token uttered by different speaker). In line with previous research (cf. Szmrecsanyi 2006; Gries 2005), the former was found to exert a stronger effect than the latter.[23] Altogether, these results reveal a very clear morphosyntactic persistence effect for the pattern's aspectual encoding.

Regarding the interaction of corpus and verb, the model shows that progressive *say* has become more frequent relative to its simple aspect competitor *I just say/said*. However, the increase fails to reach statistical significance. Importantly, much higher progressive rates than for *say* can be observed for *try* (cf. example 6.26), which clearly disfavours the simple aspect – even though the latter has become (non-significantly) more frequent in recent years. Progressive use of the second activity/event verb *go* was on the same level as *say* in BNC94 but – like *try* – shows signs of decreasing progressive affinity. The two cognition verbs *think* and *wonder* behave quite differently: While *wonder* shows a comparatively high and significantly increasing progressive affinity, *think* (increasingly) prefers use in the simple aspect pattern. In the face of *I BE just thinking*'s pronounced increase pmw, this is a surprising finding (see discussion below). As for *I BE just saying*, we can conclude that its rate of progressive encoding is lower than the one of *I'm not saying*, even tough it has increased in the past twenty years.

Finally, the question of *I BE just saying*'s productivity shall briefly be addressed. Can functionally comparable uses be found that differ in the choice of verb? Proceeding in the same way as for *I'm not saying*, uses with other communication verbs were analysed. Modestly frequent are uses with *talk, ask* and *tell*.[24] While they can occur in different contexts, it is possible to find present progressive examples that are very similar to the interpretative/argumentative uses with *say* (cf. example 6.28-30):

(6.35) *and I'm not talking about big deep drifts* ***I'm just talking*** *about say just a little dusting of maybe.* (BNC14_SCPC_1578)

(6.36) *no no no* ***I'm just I'm just asking*** *her t- which one she does she wants to do.* (BNC14_SBB2_1707)

(6.37) *it's not advice* ***I'm just telling*** *him he's wrong.* (BNC14_SE4X_124)

Analogous uses in the past tense (6.38) as well as pure recentness past tense uses (6.39) can also be identified:

[23] All effects are visualised in Figure A.14.

[24] Due to tagging inconsistencies in BNC94, present and past tense realisations with *tell* do not show up among the twenty most frequent realisations. Two tokens of *I'm just telling* and three tokens of *I was just telling* were missed by the search. While such minor inaccuracies are undesirable, they are no major cause for concern since no significance levels or overall trends were affected.

(6.38) *S0104: what? S0167: nothing* ***I was just talking*** *to my sister.* (BNC14_SE68_1758)

(6.39) *actually you know that's that's a very good point because the guy that* ***I was just telling*** *you about [...].* (BNC14_S8CB_215)

Other communication verbs such as *suggest, claim, imply* or *propose* do not occur with the *I BE just -ing* pattern, neither in BNC94 nor in BNC14.[25]

I BE just thinking/wondering The pattern's second most frequent verbal context is *think*. It shows a very pronounced and significant frequency development both in the present and the past tense. Interestingly, the latter is the more frequent variant. *I was just thinking* occurs 138 times in BNC14, compared to 91 occurrences of *I'm just thinking*. The increasing use with *think* is in line with the results of chapter 5, where the verb was identified as distinctive of present-day Face-to-face Conversation (only very closely missing statistical significance) and where an upward trend of progressive use with cognition verbs in the very same genre was reported.

Use of *I was just thinking* is exemplified in example 6.27 and 6.40-41. Similar to *say*, past progressive *think* invokes a sense of recency and is also used as part of an explanation, whereby the speaker offers an insight in his or her reasoning. The pattern functions as hedge of the actual proposition, rendering it more polite/less direct by justifying it. Like the past tense uses of *say* (example 6.31-34), past tense *think* is no obligatory progressive context from an aspectual point of view, because it does not refer to a dynamic situation in progress at a past time. Instead, we are again dealing with non-aspectual progressive use, which can additionally convey a sense of recency.

(6.40) *S0179: so it's nothing like say in athletics where you don't wanna train before you [...] compete S0058: I'm as I said it* ***I was just thinking*** *that (.) like it seems like an odd thing to do (.) but she w- obviously found it worked [...].* (BNC14_SVH7_23)

(6.41) *but* ***I was just thinking*** *that you know we don't need to be spending this kinda money on advertising for the charities [...].* (BNC14_S5XD_203)

Present tense use of the pattern is less frequent but has increased even more rapidly (+135%). It is found in contexts as the following, where it functions as a hedge and focuses on the speaker's present state of mind:

(6.42) *I don't know but* ***I'm just thinking*** *is that a way they make their money?* (BNC14_SKJ6_41)

[25] Apart from different verbal contexts, another syntagmatic context should be mentioned: The pattern *all I'm saying* seems to serve the same discourse-pragmatic function as *I'm just saying*. While it is very infrequent in BNC94 (4 tokens; 0.8 tokens pmw), it is more frequently found in BNC14 (21 tokens; 1.8 tokens pmw).

(6.43) *no I I'd be I'd be okay with that and* ***I'm just thinking*** *erm we'll just have to see [...].* (BNC14_SA2J_1723)

(6.44) *S0638: and* ***I'm just thinking*** *if I was a manager and I'd been asked to make redundancies within the team the first thing I'm gonna do is [...].* (BNC14_SBKN_2663)

Since the focus is on the present moment, progressive use of the cognition verb *think* is conceivable, even though simple aspect encoding would – from a purely aspectual perspective – be the default option; at least if one does not argue for a purely dynamic construal of *think* as a mental activity. Irrespective of the exact aspectucal nature, though, the main function of the pattern is that of a conversational hedge, whereby the speaker explains his/her point of view and – as a consequence – renders the following statement less definite.

Similar to what has been argued in the context of *say*, I suggest that the pattern's meaning should be viewed as party compositional. Regarding the focus on the present moment, continuities with the progressive's core meaning are obvious. The pragmatic side of meaning, however, results from repeated use in the appropriate contexts and does not fully emerge if the pattern's component parts are studied in isolation.

As the logistic regression model has shown (Table 6.2), the pattern's use with *think* has become less frequent relative to the corresponding simple aspect use *I just think/thought*. This means that even though the progressive pattern has become much more frequent in the speech community in the past twenty years, its simple aspect competitor must have increased even more. Compared to the other verbal realisations that went into the model, *think* is the one with the lowest progressive probability (cf. Figure A.14). Its decrease between BNC94 and BNC14 hints towards highly increased use of *think* in spoken conversation in general.

Last but not least, it is worth considering another verbal context, which is closely related to *think*: *I BE just wondering* is the fourth most frequent use of the pattern, showing a slight but non-significant upward trend (pmw) in the present and the past tense (cf. Figure 6.2 & A.12).[26] Its likelihood of progressive compared to simple aspect encoding has significantly increased between BNC94 and BNC14 (cf. Table 6.2 & Figure A.14). Frequent use of the pattern has already been noted by Römer (2005: 126,150), who argues that it serves a "politeness or softening" function. In the analysis of individual progressive verbs in chapter 5.2.2 (cf. example 5.48-50), first person uses of *wonder* were characterised as a feature of (inter-)subjective style, serving a hedging function that presents the following statement in a less definite way. Consider the following uses:

(6.45) *[...] but* ***I'm just wondering*** *why they didn't come and clear the tray away [...]* (BNC14_SNNG_1488)

[26] If measured relative to the other verbal contexts of *I BE just -ing* (i.e. not pmw), uses with *wonder* show a modest frequency decrease (cf. Figure A.13). Thus, they were slightly more typical of the pattern in BNC94 than they are today. This is in line with the findings of the Distinctive Collexeme Analysis for Face-to-face Conversation (cf. Table 5.5) in chapter 5.2.2, which revealed progressive uses of *wonder* to be typical of the 1990s data (ICE-GB).

(6.46) *[...]* ***I'm just wondering*** *if there's anything that we could do now [...].* (BNC14_SZ94_26)

(6.47) *[...]* ***I was just wondering*** *if you could email me and I can come and collect it [...].* (BNC14_SR96_227)

The politeness function is very prominent in example (6.47), where the pattern hedges a request, letting it appear less forceful. In example 6.45-46, *I'm just wondering* is used immediately before a question, highlighting the speaker's subjective perspective. As has been made clear in chapter 5.2.2, it is hard to decide whether progressive *wonder* really refers to a dynamic situation (wondering as actively reflecting) or if the stative element (reference to a temporary state of mind) is more prominent. It is argued here that from an aspectual point of view progressive encoding is not obligatory but acceptable. However, similar to what was proposed in the context of the pattern's use with *think*, it is the non-aspectual, subjective meaning component of the progressive that seems most important and that results in the pattern's discourse-pragmatic meaning as a hedge with politeness/softening function.

Assessing routinisation

Which conclusions can be drawn regarding *I BE just -ing*'s degree of routinisation? First, two of its uses – *say* (present tense) and *think* (present and past tense) – have rapidly increased in frequency (normalised pmw) over the past twenty years. They have become more strongly established in the speech community, and their status as discursive routines has clearly been strengthened. As a result, the pattern's frequency distribution has become much more skewed – a development that might be indicative of ongoing constructionalization.[27] Second, if the relative frequencies of all different lexical-grammatical realisations are considered, only one emerges that has become significantly more frequent: *I'm just saying*. All other relative frequency changes fail to reach statistical significance. From a constructionist perspective, this indicates that present tense use with *say* has become more typical of the *I BE just -ing* pattern. Third, the logistic regression model has revealed that two verbal contexts show significant changes in their aspectual preferences. While *think* – despite its frequency increase pmw – exhibits a weaker progressive affinity in BNC14 than in BNC94, uses of *wonder* have developed in the opposite direction and disfavour simple aspect use (i.e. *I just wonder(ed)*) in BNC14. Additionally, the model has shown that aspectual encoding is not only influenced by corpus and verb but is also subject to priming and recency (i.e. persistence) effects. Again, this clearly indicates that routinisation processes are not just at work on the macro-level of the community but also on the micro-level of the conversation, where transient preferences emerge and fade away.

The qualitative part of the analysis has highlighted discourse-pragmatic properties of certain verbal realisations. As predicted by Hypothesis 3, repeated use in specific contexts does indeed result in stable pragmatic associations between a syntagmatic string

[27] Cf. Goldberg et al. (2004) on the role of high token frequencies in constructional learning.

and a discursive environment (cf. Schmid 2014). Depending on verb and tense, different shades of meaning have been identified, some of which are more and others less compositional. Since all non-compositional aspects of meaning have to be learned, progressive uses such as *I'm just saying* or *I was just thinking* must be part of speakers' conversational grammars. The two cognition verbs *wonder* and *think* can be used in similar contexts, and pragmatically motivated uses with communication verbs are not restricted to *say* but can be extended to verbs such as *talk* or *tell* (cf. example 6.35-39).

Additionally, the corpus data yields a number of reduced forms of *I'm just saying* that are used as syntactically autonomous discourse markers. Their occurrence indicates that *I'm just saying* has not just become a highly routinised discursive pattern, but that – as a result of its frequent use – it has become reduced (Bybee 2015: 42f.,124f.) and finally developed into a construction of its own. Recent work by Kiesling (2018) supports this claim: He argues that *just sayin(g)* is used as a way of levelling criticism and of detaching oneself from the critical force of a previous utterance. The following conversational excerpt from BNC14 serves as an example:

S0246: what about –ANONnameM? he's single now

S0249: no thank you

S0246: no?

S0249: I thought he'd have hung himself by now

S0245: that's so hurtful –ANONnameF

S0246: that's so two thousand and eight –ANONnameF get over it

S0249: **just saying** (BNC14_S5LP)

Reduced forms of the pattern can also be found with *think* and *wonder* (*just thinking/wondering*). In most cases, however, they lack the syntactic autonomy that can be attested for the discourse marker uses of *just saying*, meaning that they are functionally equivalent to the non-reduced form.

In the face of the presented evidence, we can conclude that several aspects of *I BE just -ing*'s use and development are indicative of routinisation and pragmatic strengthening. Furthermore, the pattern's frequency increase is in line with the progressive's positive overall development in BNC94 and BNC14. The most pronounced changes pmw are observed for present progressive uses (*I'm just saying* (+161%) and *thinking* (+135%)), i.e. the most prototypical morphosyntactic contexts of the construction. From a functional perspective, however, non-prototypical aspects of use are most salient: Different verbal realisations have developed pragmatic meanings that are not primarily motivated by the progressive's aspectual semantics but by its interpretative, politeness, and/or recentness function. Most likely, non-aspectual progressive meanings take hold in the speech community via such lexical-grammatical routines. If innovative, non-prototypical uses of a construction are "tied" to specific patterns, they become predictable and can consistently be used by different speakers. In a next step, these uses can be extended to analogous contexts (e.g. semantically related verbs), becoming more and more productive and routinised.

6.3.4. *what PP BE -ing*

The most frequent of the four progressive patterns studied in this chapter is *what PP BE -ing* (cf. example 6.48-49). Since it is also the most variable one (neither tense nor personal pronoun are specified), this fact is not all too surprising. In BNC94, it reaches a frequency of 207 tokens pmw (raw freq = 1,039), which significantly increases by ca. 37% to 283 tokens pmw in BNC14 (raw freq = 3,234).[28]

(6.48) *S0179: I du n no (.) I I just have an intense dislike for Stephen Fry S0058: I know (.) that's* ***what I'm saying*** *(.) and that's fine but don't let that spill over into other people who have nothing to do with him [...].* (BNC14_SM6B_1204)

(6.49) *erm (.) and basically* ***what they're doing*** *is they're just merging the teams.* (BNC14_SECS_539)

Frequency profile

The development of the 20 most frequent lexical-grammatical realisations is depicted in Figure 6.3 (and Table A.7).[29] Also this time, the most frequent verbal context is

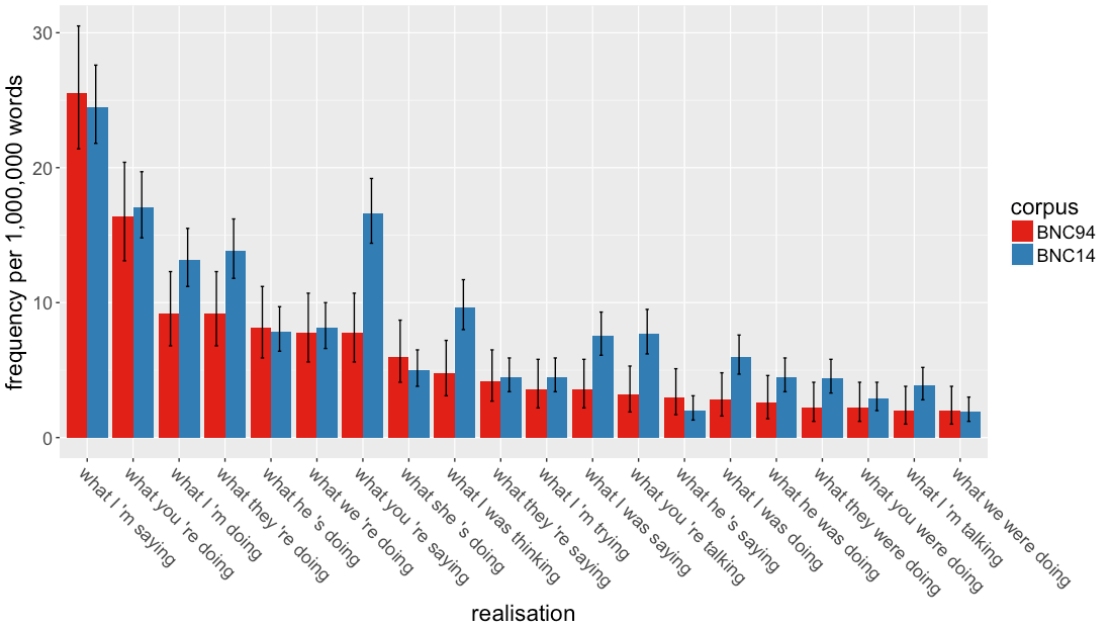

Figure 6.3.: Frequency profile of the *what PP BE -ing* pattern (20 most freq. realisations with 95% confidence intervals) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

say. More specifically, it is present tense, first person singular *what I'm saying*. Its use

[28] The change is significant for $p<0.001$.
[29] In descending order of the BNC94 frequencies.

is almost on the same level in BNC94 and BNC14, showing no significant trend pmw. Counter to that, many other of the pattern's realisations exhibit significant frequency changes. All of these changes are upward developments and not one significant downward trend has occurred. This is different, however, if relative frequencies are examined (cf. Figure A.15): *what I'm saying* significantly loses ground and only few realisations remain that show a significant upward development (e.g. *what you're saying* and *talking*). Thus, depending on whether normalised (pmw) or relative frequencies are considered, the pattern's recent development occurs in quite a different light.

The most frequent verb type among the 20 most frequent realisations is *do* (11), followed by *say* (5). Apart from these two verbs, only *talk* (2), *try* (1) and *think* (1) occur among the top twenty, which means that very few verbs account for the bulk of *what PP BE -ing*'s corpus tokens.

Specific realisations

what PP BE saying/talking As has already been pointed out, the pattern's most frequent realisation in BNC94 and BNC14 is present tense, first person singular *say* (example 6.48) (BNC94: 128 tokens, 26 tokens pmw; BNC14: 280 tokens, 25 tokens pmw). While its level of use has remained practically stable, the corresponding past tense realisation *what I was saying* has more than doubled – however, coming from a much lower level.[30]

Previous publications have identified *what I'm saying* as a recurrent pattern in spoken discourse. Simpson-Vlach and Ellis (2010: 501) mention it as a frequently used formula in academic speech, serving a meta-communicative, discourse organising function. Similarly, Biber et al. (1999: 1006) report the longer *that's what I'm saying* string – a demonstrative cleft (cf. Calude 2009) – as a common lexical bundle in conversation. As the BNC concordances reveal, the pattern is indeed frequently preceded by *that's* but also frequently followed by *is* – forming a pseudo-cleft. Consider example 6.48 as well as the following ones:

(6.50) *man you guys are hot you're fucking hot that's* ***what I'm saying****.* (BNC14_SBM6_3709)

(6.51) *yeah but I mean* ***what I'm saying*** *is when you're in a thirty mile and hour limit does it warn you?* (BNC14_S7TT_2418)

(6.52) *but I I guess* ***what I'm saying*** *is I'm against that kind of business model.* (BNC14_SL9V_3183)

These examples nicely show the meta-communicative pragmatics of the pattern, as identified by Simpson-Vlach and Ellis (2010: 501). Similar to the use of *I'm not saying* and *I'm just saying*, *what I'm saying* is used to highlight the speaker's point of view by offering an explanation/interpretation of what he or she just said or currently thinks. This means that we have again identified a progressive pattern that derives its pragmatic

[30] In Figure A.16 & A.17 the normalised and relative developments are depicted without tense distinction.

potential from the construction's interpretative function. Since it refers to an ongoing verbal exchange, the aspectual core notions of progressivity, incompletion and limited duration can also be accounted for. While the latter aspects of the pattern's meaning directly follow from progressive use, its pragmatic meaning only arises in the appropriate contexts and is not fully compositional.

Like *I was just saying*, also *what I was saying* is frequently used to make reference to recently made statements. Its use has significantly increased from 3.6 tokens pmw (raw freq = 18) in BNC94 to 7.5 tokens pmw (raw freq = 86) in BNC14 and is demonstrated in the following two examples:

(6.53) *S0162 [...] for some people it's not an easy thing to go to a developing country S0018: well that's* ***what I was saying*** *about this job in Thailand.* (BNC14_S4RF_71)

(6.54) *no that's exactly* ***what I was saying*** *twenty minutes ago.* (BNC14_SUHT_256)

The speaker stresses his/her point of view by referring back to what he/she said before – very much like the present tense use of the pattern. However, the fact that the progressive is used to refer to a completed utterance cannot be explained semantically (i.e. aspectually), but only by means of the past progressive's recentness function, which is pragmatically motivated.

The second communication verb that occurs among *what PP BE -ing*'s 20 most frequent realisations is *talk*. First person singular, present progressive use (non-significantly) increases from 2 tokens pmw (raw freq = 10) to 3.9 tokens pmw (raw freq = 44). The corresponding past tense encoding does not make it among the top twenty but is also attested in BNC94 (raw freq = 7, 1.4 tokens pmw) and BNC14 (raw freq = 15, 1.3 tokens pmw). *What I'm/was talking* is almost always followed by *about*, which means that we are actually dealing with a five-word pattern. Its present tense use has been identified as a recurrent spoken routine by Simpson-Vlach and Ellis (2010: 501).

(6.55) *yeah yeah that's what I'm saying that's* ***what I'm talking about****.* (BNC14_S8RC_1014)

(6.56) *yeah no no no I know no no no I* ***what I'm talking about*** *is you know the space where –ANONname has got the last of his stuff.* (BNC14_SUGJ_755)

(6.57) *[...] I'm doing ah something over here that actually links in to* ***what I was talking about*** *last week.* (BNC14_SHXF_157)

The pattern's meaning is closely related to the uses with *say*, which is highlighted by the fact that in example 6.55 the two verbs occur immediately after one another, referring to the same utterance. Also in example 6.56, *what I'm talking about* is used by the speaker to explain what he/she really means. The past tense example in 6.57 qualifies as recentness progressive, similar to the uses with *say* in example 6.53-54 (one could argue, however, that the explanatory/interpretative meaning is missed). As these uses with *talk* clearly show, the pattern's pragmatics as meta-communicative device are

not restricted to *say* but can productively be extended to similar verbal contexts. Again, a finding that is in line with the results for *I'm not saying* and *I'm just saying*.

As Figure 6.3 and Table A.7 indicate, second person, present tense uses with *say* and *talk* have both become much more frequent in the course of the past twenty years. Use of *what you're saying* has significantly increased from 8 tokens pmw (raw freq = 39) to 17 tokens pmw (raw freq = 190). The corresponding use with *talk* has increased from 3 tokens pmw (raw freq = 16) to 8 tokens pmw (raw freq = 88).[31]

(6.58) *I can see* ***what you're saying*** *I mean it would be an absolutely enormous problem [...].* (BNC14_SLMB_236)

(6.59) *so* ***what you're saying*** *is the the new investment needs to be not just in refurbishing and building new prisons [...] but also building housing.* (BNC14_SVXP_544)

(6.60) *[...] I have no idea* ***what you're talking about****.* (BNC14_SUC7_108)

Use of *what you're saying* in example 6.58 and 6.59 is analogous to the first person uses in 6.50-52: The speaker refers back to and interprets a previously made statement by the interlocutor. Instead of *what you're saying*, he or she could just as well have used the phrase *what you mean* – a fact that highlights the interpretative function of the pattern. Exactly the same reading applies to the use with *talking about* in example 6.60.

The other present tense uses of *say* that occur among the pattern's most frequent realisations (with personal pronouns *he* and *they*) are used in similar discourse-pragmatic contexts but remain on a low level.

Let us now address the question of *say* and *talk*'s aspectual preferences in the pattern context. To what degree do they (dis-)favour use in the corresponding simple aspect pattern (*what PP say/talk*), and how have these preferences developed over the past twenty years? As in the previous two subchapters, a logistic regression model was fitted, including the "classic" predictors corpus and verb, as well as the three persistence-related variables (previous aspectual encoding, distance, previous speaker).[32] Apart from *say* and *talk*, *do*, *try* and *think* were included in the model. For the sake of comprehensibility, both tense and pronoun distinction were not taken into account. The results are summarised in Table 6.3.[33][34]

The model reveals that *say*'s likelihood of being used in the progressive *what PP BE -ing* as opposed to the simple *what PP V* pattern has significantly increased between BNC94 and BNC14 – from roughly 30% to almost 50% (cf. Figure A.18 for the probabilities). No such trend can be observed for the uses with *talk*, whose progressive level

[31] The increase for both uses is even more pronounced if the present and past tense tense distinction is not taken into account (cf. Figure A.16).

[32] Implemented as glm(realisation ~ corp*lemma + prev_realisation*log_distance + prev_realisation*prev_speaker).

[33] All effects are visualised in Figure A.18.

[34] Note that also this model reaches a relatively small – but still acceptable (cf. chapter 4.4.5) – pseudo R^2 value (0.20).

	Dependent variable: realisation (simple vs. prog)			
	B (SE)	lower	*Odds ratio* incl. 95% CI	upper
Intercept	−2.28 (0.24) ***	0.23	0.31	0.42
Corpus and verb (pairwise comparison):				
BNC94(say) vs. BNC14(say)	0.86 (0.10) ***	1.94	2.37	2.89
BNC94(talk) vs. BNC14(talk)	0.02 (0.65)	0.28	1.02	3.65
BNC94(do) vs. BNC14(do)	0.17 (0.10)	0.98	1.18	1.43
BNC94(try) vs. BNC14(try)	1.06 (0.56)	0.97	2.88	8.58
BNC94(think) vs. BNC14(think)	0.49 (0.22) *	1.06	1.64	2.53
Persistence:				
prev_realisation(prog)	2.10 (0.25) ***	5.07	8.20	13.33
prev_realisation(prog)*log_distance	−0.30 (0.04) ***	0.69	0.74	0.80
prev_realisation(prog)*prev_speaker(same)	0.45 (0.13) ***	1.23	1.57	2.00
Model $\chi^2 = 1063.36$ ***				
Nagelkerke's $R^2 = 0.20$				
N = 6,607				

Table 6.3.: Binary logistic regression: Aspectual encoding of *what PP BE -ing* (*say, talk, do, try, think*) (*** p<0.001, ** p<0.01, * p<0.05).

has remained virtually stable. It has to be noted, however, that this level was already extremely high in BNC94 (>90%), leaving little room for further increase. Thus, uses with *say* have become more progressivised, while uses with *talk* already showed a very strong progressive preference in the 1990s, which is still evident in present-day conversation. No significant changes have occurred for *do* and *try*, whereas *think* increasingly favours progressive use (see discussion below).

Additionally, all of the persistence-related variables make a significant contribution to the model. If the previous realisation of the pattern was encoded in the progressive ("prev_realisation(prog)"), the odds of the current token to occur with the progressive are increased by more than factor eight – a syntactic priming effect. This effect interacts with the variable "prev_speaker": If the previous speaker was the same as the current one, the priming effect is stronger than if a different speaker produced the previous token. This means that we see a significant difference between production-to-production and comprehension-to-production priming. Finally, the priming effect fades away as textual distance increases. As in the previous two models, also this effect turns out as significant. These results are perfectly in line with Szmrecsanyi (2005 & 2006) and again show that aspectual choice is not just a matter of verb and corpus but also of transient psycholinguistic processes operating on the level of the conversation.

what PP BE doing/trying Let us now turn to *do*, the verb that accounts for the majority of tokens among the twenty most frequent verbal realisations of *what PP BE*

-ing. It combines with the past and the present progressive and occurs with first, second and third person pronouns. While some of its lexical-grammatical realisations remain stable (e.g. the most frequent one, *what you're doing*), others – such as *what I'm doing* – have become significantly more frequent pmw. However, if relative frequencies are considered, not one of the pattern's eleven different uses with *do* shows a significant increase. The only significant development is a downward one: *what she's doing* declines by almost 40%. While uses with *do* account for a high share of the pattern's tokens, they have not become more typical over the past twenty years.

The previous section has shown that uses with *say* serve a meta-communicative, pragmatic function. Since *do* is an activity/event verb, its uses with *what PP BE -ing* must occur in different functional contexts. Nonetheless, its pervasiveness among the pattern's most frequent realisations might indicate that also uses with *do* have acquired specific pragmatic associations. Consider the following examples:

(6.61) *I wonder* ***what he's doing*** *now he was kind of strange wasn't he?* (BNC14_SP5U_478)

(6.62) *we're gonna have to think of* ***what we're doing*** *for dinner then.* (BNC14_S5B4_886)

(6.63) *S0190: do you know* ***what you're doing****? S0227: er it's making more sense I don't know totally what I'm doing but [...].* (BNC14_SAUR_1741)

(6.64) *but* ***what he's doing*** *is he's using you (.) as an excuse to dump her (.) cos he's too much of a coward.* (BNC14_SKGU_513)

The progressive in example 6.61 qualifies as aspectual, with *now* highlighting the focus on the present moment.[35] Counter to that, the use in example 6.62 is an instance of the futurate progressive, referring to a personal plan in the near future. Still different is example 6.63: The question *do you know what you're doing?* must not be taken literally, but actually means *do you know how to do it?*, implying a sense of doubt. Finally, the progressive in example 6.64 (as well as the one in 6.49) is part of the speaker's interpretation of a situation. Of all four examples, this interpretative use is clearly most similar to the meta-communicative uses with *say* and *talk*. As these examples demonstrate, the pattern's uses with *do* are very variable – an impression that is further supported by qualitative analysis of a larger number of corpus tokens. While *what PP BE doing* can certainly be used pragmatically (example 6.63-64), this is not consistently the case (example 6.61-62). Moreover, also the pragmatically motivated uses are not necessarily interpretative in nature, as is highlighted by example 6.63.

Apart from *do*, only one further activity/event verb (*try*) occurs among the pattern's 20 most frequent realisations. Use of *what I'm trying* is modestly frequent in BNC94 (raw freq = 18, 3.6 tokens pmw) and BNC14 (raw freq = 51, 4.5 tokens pmw), showing no significant frequency development. It is virtually always followed by a *to*-infinitive,

[35] More specifically, it refers to (temporary) habitual behaviour applying at the present moment.

making it a five-word pattern. Particularly frequent verbs in the infinitive position are *say* and *do*, as in the following two examples (note also preceding *that's*):

(6.65) *yeah but that's* ***what I'm trying to*** *say Subway is probably your best option.* (BNC14_SL9V_838)

(6.66) *[...] I think that's* ***what I'm trying to*** *do with the applications I'm writing at the moment [...].* (BNC14_SRFV_1294)

In example 6.65, the pattern serves a meta-communicative function like the one attested for *say*. Instead of *what I'm trying to say*, the speaker could just as well have said *what I'm saying*. Use of the activity/event verb *try* only masks meta-communicative use of the pattern. In the second example, the infinitive is formed with *do*. Again, an interpretative reading seems to be present, in the sense that the speaker reflects on/interprets his or her current behaviour (which is further supported by the discourse marker *I think*).

Regarding the aspectual preferences, the regression model (Table 6.3) reveals that neither uses with *do* nor *try* have significantly shifted towards the progressive in the course of the past twenty years. For both verbs, the odds of occurring with progressive *what PP BE -ing* as opposed to simple *what PP V* have only slightly increased. As Figure A.18 shows, *try*'s probability to occur with the progressive is very high (ca. 90% in BNC14), roughly on the same level as *talk*. Counter to that, the level for *do* is much lower (ca. 40% in both corpora), meaning that its use is much less progressivised than the one of *try*.

what PP BE thinking As in the *I BE just -ing* pattern, *think* also occurs among the most frequent realisations of *what PP BE -ing*. The first person singular, past tense use *what I was thinking* is found ca. five times pmw in BNC94 (raw freq = 24 tokens) and significantly[36] increases to 10 tokens pmw (raw freq = 110) in BNC14.[37] If the tense distinction is left aside and also the present progressive uses are taken into account, the increase is even more pronounced, resulting in a BNC14 frequency of about 15 tokens pmw (cf. Figure A.16).[38] The following examples highlight typical contexts of use:

(6.67) *S0441: did he not could he not have sent a message either? S0439: that's* ***what I was thinking*** *I was like I did try and call and ring people but the signal at –ANONplace it's always so horrendous that I'm most of the time I think no one ever gets your message [...].* (BNC14_S2AJ_29)

(6.68) *S0097: yeah (.) he'd have probably organised a meeting or something so he didn't have to deal with it S0151: that's* ***what I was thinking*** *(.) I was surprised he was there actually.* (BNC14_S66A_851)

[36] The increase is significant for $p<0.01$.

[37] Note, however, that the increase fails to reach significance if it is calculated relative to the other uses of the pattern and not pmw.

[38] Since present progressive *what I'm thinking* very closely fails to make it among the pattern's 20 most frequent realisations in BNC94 (1.8 tokens pmw), its frequency increase in BNC14 (5.3 tokens pmw) is not depicted in Figure 6.3 & A.15.

(6.69) *S0084: so it's what it's how you get satisfaction from your job I suppose S0041: that's* ***what I'm thinking*** *(.) and am I really gonna get satisfaction out of doing that?* (BNC14_SXCW_616)

All three examples of *what I BE thinking* are preceded by *that's* – the pattern's preferred context both in BNC94 and BNC14. In the former, 50% of all tokens co-occur with it, compared to 61% in the latter. Thus, it can be argued that we are actually dealing with a six-word pattern.

In example 6.67-69, it functions as an emphatic marker of speaker agreement, occurring at the beginning of a turn as an immediate response to the interlocutor's statement. Furthermore, the past tense realisations are frequently used to agree with assumptions that eventually turned out as wrong, juxtaposing one's expectation with the real course of events. This reading is particularly salient in example 6.68. Regarding the aspectual nature of the situation, the progressive is chosen because it highlights the speaker's state of mind at the present moment (present progressive) or at a particular moment in the past (past progressive). However, since *think* does not refer to a clearly dynamic reflection process but to a state-like point of view, this momentary focus does not make progressive encoding obligatory. In all three examples, simple aspect encoding would – at least from a purely semantic perspective – be the default. Pragmatically, the progressive functions as foregrounding device, highlighting speaker agreement by explicitly focusing on the speaker's present/past state of mind, i.e. his or her subjective opinion. The only meaning aspect that is completely unrelated to the progressive and that does neither follow from the other component parts of the pattern is the supposed juxtaposition function of the past tense uses. Its routinisation fully depends on repeated usage in the appropriate contexts.

The regression model (Table 6.3) has revealed that the odds of *think* to occur in the progressive *what PP BE -ing* as opposed to the corresponding simple aspect *what PP V* pattern have increased by factor 1.64 between BNC94 and BNC14. Although this is no dramatic shift towards progressive encoding, it is statistically significant. As Figure A.18 shows, *think* has the weakest progressive affinity of the five verbs that went into the model, both in BNC94 and in BNC14. Since cognition verbs are not normally considered typical progressive verbs, this is not very surprising. It means that even though in this particular pattern context *think*'s progressive encoding has become more frequent, it is still the corresponding simple aspect encoding (e.g. *what I thought*) that accounts for the majority of uses.

Use with cognition verbs other than *think* is rare in both corpora. Nonetheless, a number of tokens very similar to the ones in example 6.67-69 can be found, suggesting that the *what PP BE thinking* pattern can productively be expanded to related verbal contexts. Consider the following two BNC14 hits:

(6.70) *S0589: maybe the spores were already on it S0588: yeah that's* ***what I'm wondering*** *(.) but I I didn't notice when I moved it from downstairs but let's check [...].* (BNC14_S7NJ_320)

(6.71) *S0530: that looks like eighteen S0529: I know that's* ***what I'm hoping*** *but it could be twelve as well.* (BNC14_S954_496)

Assessing routinisation

The already high frequency of *what PP BE -ing* in BNC94 has further increased in BNC14, meaning that the pattern's status as a discursive routine has been consolidated in the course of the past twenty years.

The verbal paradigm is dominated by *do* and *say*, which account for the majority of the 20 most frequent lexical-grammatical realisations. Similar to the previous two patters, uses with *say* serve a meta-communicative function that shows continuities with the progressive's interpretative meaning. Additionally, if used with the past tense, progressive *say* often refers to recent communicative events. In relative terms, the pattern's most frequent use *what I'm saying* has significantly lost ground, whereas the corresponding second person use *what you're saying* has increased its share. Uses with *do* are functionally very variable, showing no dominant pragmatic meaning. Also frequency-wise, no clear developmental trends can be discerned. Of the remaining verbs, *think* is particularly interesting, because its first person singular uses have rapidly increased and frequently function as pragmatically salient agreement markers. Also use of *what you're talking about* has become much more frequent.

In most cases, the identified pragmatic functions – i.e. interpretation/explanation, recentness, emphasis/agreement – can readily be inferred from the discursive context, meaning that they are not highly opaque. While the present progressive uses almost always show continuities with the construction's aspectual semantics, this is not so evident for the past progressive uses referring to recent but completed communicative events.

With regard to the aspectual preferences, the regression model has revealed (significantly) increasing progressivisation of *say* and *think*. Furthermore, it has shown a particularly strong progressive bias of *talk* and *try*. Like the previous two models, also the present model confirms that aspectual choice is further influenced by previous aspectual encoding (priming) and by textual distance (recency).

These results show that several lexical-grammatical realisations of *what PP BE -ing* are well-established conversational routines, which are frequently used as meta-communicative devices – similar to *I'm not -ing* and *I BE just -ing*. The pattern's continuing frequency development indicates ongoing entrenchment in the speech community.

6.3.5. *what BE you -ing*

The final progressive pattern to be studied is the question *what BE you -ing*. Diachronically, its usage remains relatively stable, with a frequency of 93 tokens pmw (raw freq = 466) in BNC94 and 96 tokens pmw (raw freq = 1,099) in BNC14.[39] By far the most frequent lexical-grammatical realisation is second person, present progressive use with *do* (example 6.72), followed by the equivalent use with *talk* (example 6.73).

[39] The change fails to reach statistical significance.

(6.72) *I was like* ***what are you doing****? why are you doing this to me?* (BNC14_S8PT_812)

(6.73) *I got my flight for four hundred quid (.) and he was like* ***what are you talking about****? and I said oh I got my flight in the January sale.* (BNC14_SU82_3473)

Frequency profile

Figure 6.4 and Table A.8 depict the frequency development of the pattern's 15 most frequent realisations.[40][41]

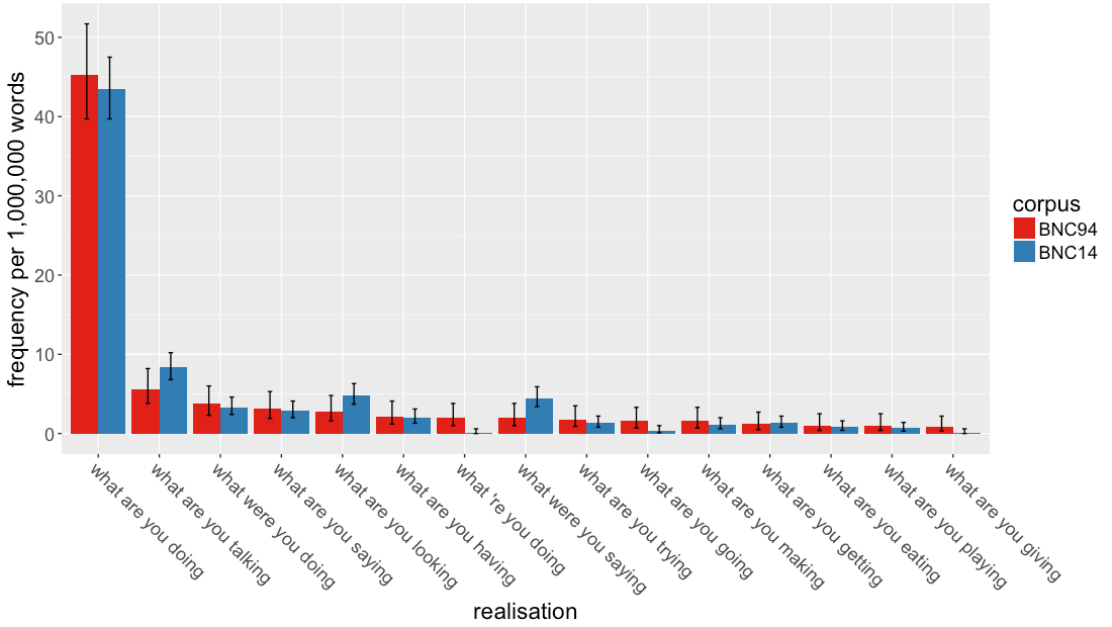

Figure 6.4.: Frequency profile of the *what BE you -ing* pattern (15 most freq. realisations with 95% confidence intervals) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

As has already been pointed out, *what are you doing* is by far the most common use. It shows no significant frequency change between BNC94 and BNC14, reaching a level well above 40 tokens pmw in both corpora. This makes it the most frequent realisation not only of *what BE you -ing* but of all four patterns studied in this chapter. As Figure 6.4 also shows, a few contracted uses of *what are you doing* are attested in BNC94 (10 tokens). However, compared to the realisations with un-contracted *are*, the numbers are marginal.[42]

[40] In descending order of the BNC94 frequencies.

[41] Since most of the pattern's lexical-grammatical realisations are very infrequent, I decided to report only the 15 most frequent uses.

[42] *what are you doing* is the only realisation for which a number of contracted uses are attested. For all

The pattern's second most frequent use *what are you talking* does come nowhere near the frequency of *what are you doing*. In BNC94, it occurs ca. 6 times pmw (raw freq = 28), compared to ca. 8 tokens pmw (raw freq = 95) in BNC94.[43] The only realisation showing a significant frequency increase pmw is *what were you saying*, which is attested only two times pmw in BNC94 (raw freq = 10) but 4.5 times in BNC14 (raw freq = 51).

Since the pattern's overall use is more or less stable and no dramatic changes of specific realisations have occurred, the relative frequency development (Figure A.19) is very similar to the development pmw and does not reveal major typicality shifts among the different uses.

Specific realisations

what are you doing Biber et al. (1999: 1008) identify *what are you doing* as a lexical bundle in conversation, occurring with a frequency of at least 40 tokens pmw. This is in line with the current findings for BNC94 and BNC14, where the pattern reaches a frequency of ca. 45 tokens pmw. Schmid (2014: 272) also makes reference to the pattern and argues that it is used by the speaker to inquire about the intentions of the hearer, which means that it has acquired a pragmatic function. Finally, there is Kay and Fillmore's (1999) famous paper about the *What's X doing Y?* construction, in which the authors argue that

> as a part of the grammar of English, the *What's X doing Y?* construction [...] directly encodes, in addition to a request or demand for an explanation, the pragmatic force of attributing what we call INCONGRUITY to the scene or proposition for which the explanation is required. (1999: 4)

The form *is* is not fixed but can be changed to other forms of *be* (1999: 20). Consequently, the construction can, for example, be realised as *what are you doing here?* (1999: 22).

Kay and Fillmore argue that the progressive does not convey aspectual meaning in this constructional context (1999: 6,22). Instead, it supposedly serves an emphatic/subjective function, adding to the notion of surprise and incongruity.

The following tokens and the one in example 6.72 illustrate the pattern's use in the conversational BNC data:

(6.74) *and I bumped into him and he was I was like oh my god* ***what are you doing*** *here?* (BNC14_S5YC_271)

(6.75) *and I did get a few kinda strange looks as if they say like* ***what are you doing*** *up here eh by yourself?* (BNC14_SRZT_44)

(6.76) *S0041:* ***what are you doing*** *now? S0086: em I'm getting a yoghurt.* (BNC14_SLDB_508)

other realisations, only very few contracted uses can be found. Thus, *what BE you -ing* behaves in the exact opposite way as the other three patterns, were contracted forms of *be* are the default and full forms are extremely rare.

[43] The change fails to reach statistical significance.

(6.77) ***what are you doing*** *next week? you should come to –ANONplace.*
(BNC14_SXCB_316)

The uses in example 6.72 and 6.74-75 are very much in line with Schmid's and Kay and Fillmore's analysis of the pattern's discourse-pragmatic functionality. The speaker inquires about the intentions of the hearer, simultaneously expressing surprise and/or disapproval – i.e. what Kay and Fillmore refer to as incongruity. Note that each of the three uses is preceded by a quotative – 6.72 and 6.74 by *be like* and 6.75 by *say like*. This is not the case in example 6.76, where the speaker's inquiry about the hearer's intentions is evident but where the notion of incongruity cannot unambiguously be attested. The adverb *now* enforces the progressive's aspectual meaning, invoking a strong sense of ongoingness. Even though no temporal adverbs are present in example 6.72 and 6.74-75, the situations referred to are clearly in line with the progressive's aspectual semantics. The primary focus is on an activity/event that is happening at the present moment. The pragmatic associations have developed on top of this aspectual reading but are certainly not at odds with it. Consequently, Kay and Fillmore's claim that the progressive has lost its aspectual meaning in the *What's X doing Y?* construction seems somewhat doubtful. Nonetheless, we are dealing with a partly idiomatic pattern, whose conventionalised (pragmatic) meaning is more than just the sum of the individual component parts.

As example 6.77 shows, *what are you doing* can also be used with the progressive's futurate meaning, referring to personal plans or arrangements. While these uses are comparatively common in the BNC data, they lack the pragmatics of incongruity and are functionally very much distinct from example 6.72 and 6.74-75. One could argue, however, that they constitute the logical extension of uses such as 6.76, where the line between the present moment and the immediate future is not well-defined.

Let us now focus on the question of aspectual preference. Like in the previous subchapters, a logistic regression model was fitted to the data. Unlike the previously reported models, the present model does not just focus on the verb level, but uses a combined variable of verb and tense. This is due to the fact that with present progressive *do* a single lexical-grammatical realisation exists that accounts for the vast majority of tokens of the *what BE you -ing* pattern. Thus, the model illuminates to what degree the WHAT YOU DO(present) question is biased towards the progressive and whether *what are you doing* is crowding out *what do you do*. In this context, it is important to note that the pattern's second most frequent realisation – present progressive *talk* – was not included in the model. This is because no occurrences of the corresponding simple aspect use *what do you talk* were found in the BNC94 data, meaning that there was no variation to be modelled.[44] Instead, the present and past tense uses of *say* were taken into account (see discussion below). The influence of the persistence-related variables was quantified in the same way as in the previous three models. The model output is summarised in Table 6.4.[45]

First, it can be seen that there is no stronger progressive bias for present tense *do*

[44] A word search for *what do you talk* yields one token in BNC94. However, infinitive *talk* is tagged as VVB-NN1, meaning that the token was not identified by the relevant POS-tag search.

[45] Implemented as glm(realisation ~ corp*lemma + prev_realisation*log_distance).

	Dependent variable: realisation (simple vs. prog)			
	B (SE)	lower	*Odds ratio* incl. 95% CI	upper
Intercept	−3.56 (0.48) ***	0.22	0.44	0.87
Corpus and verb (pairwise comparison):				
BNC94(do_present) vs. BNC14(do_present)	0.03 (0.21)	0.68	1.03	1.56
BNC94(do_past) vs. BNC14(do_past)	0.29 (0.42)	0.58	1.33	3.04
BNC94(say_present) vs. BNC14(say_present)	1.45 (0.50) **	1.60	4.27	11.40
BNC94(say_past) vs. BNC14(say_past)	1.66 (0.56) **	1.76	5.28	15.85
Persistence:				
prev_realisation(prog)	3.75 (0.55) ***	14.77	42.58	128.93
prev_realisation(prog)*log_distance	−0.46 (0.08) ***	0.53	0.63	0.74
Model $\chi^2 = 496.98$ ***				
Nagelkerke's $R^2 = 0.45$				
$N = 1{,}222$				

Table 6.4.: Binary logistic regression: Aspectual encoding of *what BE you -ing* (*do_present, do_past, say_present, say_past*) (*** p<0.001, ** p<0.01, * p<0.05).

("do_present") today than twenty years ago. The increase in odds is minimal and the effect far from being significant. Basically the same is true for the past tense realisations ("do_past"), whose progressive/simple ratio neither shows a significant change. However, as effect plot A.20 (Appendix) reveals, the pattern's progressive probability is much higher for present tense *do* (ca. 70%) than for the past tense use (ca. 20%). This means that progressive use is much more routinised for *what are you doing* than for *what were you doing*. In the latter case, simple aspect *what did you do* is much more common. This shows that *what are you doing* is not just by far the most frequent realisation of the *what BE you -ing* pattern, but that it also exhibits a clear and diachronically stable progressive bias.

As in the previous models, the observed variation in the data does not only go back to verb and corpus, but can partly be explained by persistence effects in spoken discourse (cf. Szmrecsanyi 2006). In the present model, two persistence-related predictors turned out as significant: First, aspectual encoding of the pattern's previous token ("prev_realisation"), which is interpreted as a syntactic priming effect. If the previous token was realised as progressive, the odds of the current token to be realised as progressive are increased by factor 42.58 (compared to previous simple aspect encoding). Second, previous aspectual encoding significantly interacts with (logarithmic) textual distance in words ("prev_realisation(prog)*log_distance"), meaning that the syntactic priming effect weakens as distance in words increases. As above, this interaction is interpreted as a recency effect of priming. The difference between production-to-production and comprehension-to-production priming turned out as predicted (the former being

stronger than the latter) but did not reach statistical significance. Thus, the relevant interaction term ("prev_realisation*prev_speaker") was excluded from the model.

what are you talking The pattern's second most frequent realisation is *what are you talking*, which is almost always followed by *about* (93% in BNC94; 98% in BNC14) – making it in effect a five-word pattern. As already explained above, its frequency increase between BNC94 and BNC14 fails to reach significance, with its level of use remaining below 10 tokens pmw. This number is in line with Biber et al. (1999: 1009), who list *what are you talking about* as a conversational lexical bundle with a frequency between five and ten occurrences pmw. Example 6.73 and the following ones show typical contexts of use:

(6.78) *she was like happy New Year* ***what are you talking about*** *it's hours away?* (BNC14_SUVQ_3444)

(6.79) *S0492: bubble bath's the same (.) it's the same thing S0493: it's not the same thing at all* ***what are you talking about****?* (BNC14_SCRM_1682)

(6.80) *which table?* ***what are you talking about****? that table?* (BNC14_SQ6Z_1494)

It is evident that these uses of *what are you talking about* must not be understood in their literal sense but that they have acquired a strong pragmatic meaning. Similar to *what are you doing*, the speaker attributes a sense of incongruity to the scene, often alongside a notion of surprise or disapproval. The pattern's meaning can be paraphrased as *what do you mean* or – at least in some cases – as *are you serious* (which have both acquired pragmatic meanings themselves). At this point, it is worth highlighting the connection to pattern III and its realisation *what you're talking about*, which is used in closely related contexts (cf. example 6.60). Finally, *what are you talking about* is frequently introduced by quotative *be like* (cf. example 6.73&78) – another parallel with the more frequent *what are you doing*.

The pattern's progressive bias is extremely pronounced. Not one simple aspect token of the form *what do you talk* was found in BNC94, meaning that there was no variance to be explained by the logistic regression model.[46] Also in BNC14, only five simple aspect tokens are attested, which amounts to a progressive probability of 95%. These high values are similar to the ones observed for *talk* in the *what PP BE -ing* pattern (cf. Figure A.18).

Thus, *what are you talking about* is a modestly frequent conversational routine with a very strong and stable progressive bias. Its frequency pmw has increased notably but non-significantly, which means that its status in the conversational grammar of English has not changed dramatically over the past twenty years. The pattern has developed a distinctive (non-compositional) pragmatic meaning, which – alongside its strong progressive bias – is indicative of a high degree of routinisation and possibly also of holistic storage.

[46] Inclusion in the model would have resulted in an extremely inflated standard error of the *B* coefficient and in very large confidence intervals in the effect plots.

what BE you saying Before concluding this subchapter, I would very briefly like to address *what BE you -ing*'s uses with *say*. While present progressive *what are you saying* shows no significant frequency development (ca. 3 tokens pmw both in BNC94 and BNC14), usage of past progressive *what were you saying* has more than doubled (2 tokens pmw in BNC94, 4.5 tokens pmw in BNC14).[47] Consider the following examples:

(6.81) *so I said* ***what are you saying****? is that how much money I'm going to be getting back then twenty-two quid?* (BNC14_SB4D_1334)

(6.82) *I've got mushrooms what do you reckon? or we could go out but it's cheaper this way isn't it ? (.)* ***what are you saying*** *–ANONname?* (BNC14_SR8V_1204)

(6.83) *can you can you speak up please?* ***what are you saying****?* (BNC14_S2JK_493)

(6.84) *yes I know it's Brownies today but –ANONname* ***what were you saying*** *earlier?* (BNC14_SQBK_5)

(6.85) *sorry my love* ***what were you saying****?* (BNC14_S4CU_387)

Arguably, present progressive use in example 6.81 is functionally similar to the use of *what are you talking about*. The speaker wants to know what the hearer actually means and seems to express his or her surprise and disapproval. Again, possible paraphrases could be *what do you mean* or *are you serious*. Counter to that, the pattern's meaning in example 6.82 is not one of surprise or disapproval but of neutral inquiry regarding a previously made suggestion. Earlier in 6.82, the speaker uses the phrase *what do you reckon*, which pretty much captures the meaning of *what are you saying* in this particular context. Still another meaning can be observed in example 6.83, where *what are you saying* does not carry any pragmatic load but is simply used in its literal sense following the request *can you speak up please*. In the light of these examples, *what are you saying* appears to be functionally more variable than *what are you talking about*, meaning that it is not consistently associated with a distinct discourse-pragmatic meaning.

If used with the past progressive (example 6.84-85), the pragmatics of incongruity can neither be discerned. Instead, *what were you saying* is used to refer to a recent, completed communicative event. Thus, it qualifies as recentness progressive, lacking the construction's imperfective aspectuality.

The logistic regression model (Table 6.4) reveals that both the pattern's present and past progressive uses with *say* have become significantly more frequent compared to the corresponding simple aspect realisations (i.e. *what do/did you say*). Effect plot A.20 shows that with a share well below 10% use of *what were you saying* is very rare in BNC94 compared to *what did you say*. In BNC14, the simple aspect realisation is still the dominant variant but the share of progressive encoding has risen to around 20%. Compared to that, the increase of present progressive *what are you saying* has occurred on a considerably higher level.

[47] The change is significant for $p<0.05$.

Assessing routinisation

What conclusions can be drawn regarding *what BE you -ing*'s degree of routinisation? First, one of its realisations – *what are you doing* – accounts for the lion's share of all tokens. Its high discourse frequency alone indicates that it must be well-anchored in speakers' conversational grammars. Since no significant frequency change has occurred, this status has remained unchanged over the past twenty years. *What are you doing* is much more frequent than its simple aspect competitor *what do you do*, both in BNC94 and BNC14, showing that its progressive encoding is well-entrenched. The pattern is frequently used in contexts in which it attributes a sense of incongruity to the scene (cf. Kay and Fillmore 1999: 4,22), expressing the speaker's surprise and/or disapproval. Thus, in addition to its high frequency of use and its progressive bias, *what are you doing* has developed a prominent pragmatic meaning, which is central to the claim that we are dealing with a firmly established conversational routine that may well have acquired construction status.

The pattern's second most frequent lexical-grammatical realisation – *what are you talking about* – is much less frequent but does nonetheless show signs of advanced routinisation. It is mainly used pragmatically to express speaker surprise and/or disapproval (i.e. incongruity), not unlike *what are you doing*. What is more, it shows a very strong progressive bias, with no or very few corresponding simple aspect encodings in BNC94 and BNC14.

Finally, *what BE you -ing*'s present progressive use with another communication verb, *say*, has been shown to be functionally very variable, exhibiting no stable pragmatic associations. The respective past progressive use *what were you saying* is gaining ground and favours use in contexts in which it refers to recent but completed communicative events, i.e. qualifying as recentness progressive.

6.4. Discussion of results

The present chapter has analysed the progressive's usage and development in spoken British English from the perspective of recurrent patterns of use. Based on previous research by Römer (2005), it was argued that the construction must not be treated as an exclusively grammatical but also as a lexical-grammatical phenomenon, with a whole range of lower-level schemas that are guiding speakers' linguistic behaviour. The progressive output from DCPSE and FCSE was used to identify the four most frequent four-word patterns, which were then further analysed in the much larger BNC94 and BNC14 corpora. As predicted by Hypothesis 3, many lexical-grammatical realisations of these patterns have developed pragmatic meanings, showing signs of advanced routinisation in the speech community. Furthermore, the analysis has revealed that certain patterns – such as *I'm just saying* – have dramatically increased in frequency over the past twenty years, while others – such as *what are you doing* – have remained stable.

The connection between pragmatic strengthening and frequency of use was introduced in chapter 2.1.2, arguing that repeated usage in specific contexts can result in pragmatically strengthened strings that might eventually acquire construction status (cf. Bybee

2006: 721f., 2007:17, 2015: 133ff.). For such a process to happen, already low frequencies can be sufficient (ibid.). This chapter has identified several contexts of progressive use that have developed exactly along this line. While some are comparatively frequent – e.g. *what are you doing* or *what I'm saying* – others are much rarer but have nonetheless acquired stable pragmatic meanings – e.g. *what are you talking about* or *I'm not being funny*.

As Schmid (2014) has shown, pragmatic strengthening (or as he calls it, the development of pragmatic associations) is pervasive among lexical-grammatical patterns and often a central aspect of their use. He lists almost 50 lexical bundles occurring in natural conversation (all identified by Biber et al. 1999: 1001ff.) and argues that all of them "can easily be associated with quite specific pragmatic functions" (2014: 273). He claims that the listed patterns indicate that

> lexical bundles are a mirror of what people in face-to-face social interactions most frequently negotiate: they exchange information concerning states of minds, intentions and plans for future actions, motivations for past actions; they reject each others' opinions and try to secure understanding; they give and ask for permission; they inform each other about their intentions, and so on. (2014: 273)

Many of the patterns studied in this chapter serve exactly such, or similar discursive functions: Meta-communicative uses as *I'm not saying*, *I'm just saying* or *what I'm/was saying* are used to secure understanding and to explain oneself. *I'm/was just wondering/thinking* introduces the speaker's state of mind/opinion in a polite way. *What are you doing* and *what are you talking about* often convey surprise and disapproval, while *I'm not having* is frequently used to indicate rejection of a particular event or state of affairs.

The analyses also revealed that it is possible to find continuities with the progressive's grammaticalized meaning in many cases. These can be more or less obvious/direct, but there are certainly not many lexical-grammatical progressive uses that defy any semantic connection to the construction's core meaning. In most cases, the attested pragmatic meaning has developed out of and eventually on top of the progressive's aspectual semantics, resulting in quite complex conventionalised meanings of the respective patterns. As has already been explained in the introductory part of the present chapter, this observation is backed by work of Bybee and Torres Cacoullos (2009), who state that "even though prefabs develop their own discourse-pragmatic characteristics, they retain traces of the constrains on their associated construction" (2009: 192).

Connecting these results to those of chapter 5, it must be asked how far the development of specific patterns is in line with the observed overall trends. Chapter 5.1 reported an increase in progressive use from approximately 644 tokens/htw in BNC94 to 759 tokens/htw in BNC14. Regarding the four studied patterns, we have seen that two have increased as well (pattern II & III), while one has remained stable (pattern IV) and one has decreased in frequency (pattern I). This shows that the progressive's overall development is not necessarily indicative of the development of frequent patterns of use. Since the studied patterns often have acquired specific functional associations and their

share among all progressive tokens is a modest one (ranging from ca. 9 tokens/htw to ca. 28 tokens/htw depending on pattern and corpus)[48], this is neither very surprising nor problematic, though.

In the course of this chapter, I have frequently made reference to the so-called recentness progressive. Past progressive patterns such as *what were you saying* or *I was just saying* often refer to recently made statements (without invoking a sense of imperfectivity). In line with these results, the HCFA in chapter 5.3 showed that combinations of the past progressive with communication verbs are overrepresented in Face-to-face Conversation in LLC, ICE-GB and FCSE, indicating that the recentness progressive is indeed a feature of conversation as claimed by Leech (2004: 32). The analyses have highlighted specific lexical-grammatical patterns that are used to convey this non-prototypical progressive meaning. Such routinised form-to-function mappings, which are associated with specific discourse-pragmatic contexts, can guide speakers' linguistic choices and make non-prototypical uses of the progressive predictable and learnable. The same applies to meta-communicative present progressive patterns such as *I'm not saying*, *I'm just saying* or *what you're saying*, whose functionality can be derived from the progressive's interpretative meaning.

It has also been shown that certain patterns involving cognition verbs – *I BE just thinking/wondering* and *what I BE thinking* – are comparatively frequent and increasing in use (esp. the uses with *think*). This is in line with the results of chapter 5.2.2, which revealed an increased level of progressive use with cognition verbs. In the light of Bybee and Torres Cacoullos's claim that prefabs can serve "as the loci for extensions of [...] construction[s]" (2009: 212), it seems conceivable that routinised progressive use with *think* and *wonder* in specific lexical-grammatical contexts helps to facilitate the overall spread of cognition verbs with the progressive.

Using frequency-based measures, different aspects of the patterns' use and development could be pointed out: While some patterns are strongly represented in the speech community – and possibly also on the mental level – (e.g. *what are you doing*), others exhibit changes regarding their typicality (e.g. *what I'm saying*), and some have developed a significantly stronger attraction to the progressive (e.g. *what were you saying* vs. *what did you say*).

In addition to that, I analysed the degree of compositionality of selected lexical-grammatical realisations. Highly idiomatic (*I'm not being funny*) as well as partly compositional (e.g. *what are you talking about* or *I'm not saying*) and fully compositional uses (e.g. *what he's doing* in example 6.61) were identified. If non-compositionality is taken as an indicator of holistic storage, most of the analysed lexical-grammatical progressive uses are potentially stored as chunks since their meaning cannot fully be derived from their component parts. As such, they would have acquired the status of cognitively real form-meaning pairings, i.e. constructions. However, in the face of lacking psycholinguistic evidence, this claim has to be treated with some caution.

[48] Note that the patterns' overall frequencies have so far been reported per million words (pmw). In order to enable easy comparability with the overall frequencies reported in chapter 5, I decided to report them per hundred thousand words (htw) at this point.

An exception in this regard is reduced *just saying*, which has developed into a syntactically autonomous discourse marker (Kiesling 2018), i.e. an independent construction. Frequent use of *I'm just saying* in appropriate contexts has led to the pattern's reduction and – finally – its emancipation from the source construction, constituting a clear case of constructionalization.

Focusing on usage in related verbal contexts, the issue of the patterns' productivity was addressed. For example, *I'm not suggesting* can be used with the same meaning as the much more frequent *I'm not saying*, indicating that the latter is – at least to a certain degree – productive. Since productivity can be regarded as an indicator of schematization (cf. chapter 2.1), it can be assumed that speakers do not only have at hand the *I'm not saying* pattern with its associated contextual information but also a more schematic *I'm not -ing*$_{communication}$ pattern, exhibiting very similar contextual associations.

The final point in the discussion of routinisation concerns the effects of priming and recency, linking to previous work by Szmrecsanyi (2006) and Gries (2005) as well as to Pickering and Garrod's (2004 & 2017) concept of interactive alignment (cf. chapter 2.1.1). The logistic regression models revealed that the choice between progressive and the respective simple aspect encoding of a pattern is not only determined by corpus and verb but also by the aspectual encoding of the pattern's previous token in discourse. This increased readiness to use the progressive version of a pattern if the previous use of the pattern was also encoded as progressive (and vice versa for the simple aspect) was interpreted as syntactic priming effect. It turned out as significant for all four patterns. The regression modelling also revealed a recency effect of priming: With increasing textual distance to the preceding token (measured as the logarithmic distance in intervening words), the priming effect weakened. These results are in line with previous corpus priming research (Gries 2005; Szmrecsanyi 2006) and indicate that interactive alignment does not only affect the emergence of transient linguistic routines but also the use of well-established ones: Pattern choice is significantly influenced by what is being uttered in an ongoing conversation. While the present analysis has only focused on aspectual encoding, it is likely that also other aspects of pattern use are subject to priming (e.g. verb (class), tense, etc.). This insight is important because it shows how both transient psycholinguistic constraints (priming effect) and stable constraints of the community grammar (effect of corpus) influence speakers' linguistic choices in dialogue.

In order to infer what this could mean for language change, one must address the question how transient priming effects can ultimately result in long-term frequency shifts (cf. Pickering and Garrod 2017; Mair 2017). Since priming effects apply to all sorts of linguistic items, what is needed is some kind of additional factor that influences selection. Regarding recurrent patterns of use, such a selective advantage could be pragmatic salience. Patterns that have acquired specific pragmatic associations are more readily selected in a given discourse situation than pragmatically neutral ones (cf. Schmid 2014: 277ff.). Thus, while all kinds of lexical-grammatical patterns can potentially be primed, it is the pragmatically salient ones that become established as conventionalised routines because they exhibit an additional way of attracting speakers' attention.

Finally, I would like to discuss the nature of the changes that have been reported in the

present chapter. Does pragmatic strengthening and routinisation of specific progressive patterns constitute constructional change, and is it indicative of ongoing grammaticalization?

We are indeed observing another case of constructional change as defined by Hilpert (2013a: 16). If we conceive of the studied patterns as subschemas of the progressive construction – either partially filled (e.g. *I BE just -ing*) or fully fixed (e.g. *I'm just saying*) – it is apparent that the observed frequency changes of some of these lower-level schemas will correspond to changes in the progressive's constructional network. What is more, relative frequency changes (e.g. increase of *what you're saying* and decrease of *what I'm saying* among the uses of *what PP BE -ing*) indicate that the network of a particular subconstruction has been altered in the course of the past twenty years. Finally, the pragmatic strengthening of specific lexical-grammatical progressive uses constitutes a functional change that will most likely be registered in the construction's larger network.

The issue of grammaticalization is less straight forward. On the one hand, increasing use in contexts in which the progressive is not obligatory from an aspectual point of view – e.g. some of the meta-communicative uses – may very well indicate that the construction's semantic constraints are gradually loosening. The same applies to past progressive patterns that are used as recentness progressives, or to present progressive patterns formed with *think* or *wonder*, which clearly do not constitute prototypical functional contexts. On the other hand, it must not be forgotten that the studied patterns only account for a fraction of all progressive tokens and that they can also be realised with activity/event verbs expressing aspectual meaning. Furthermore, the progressive's functional paradigm has always been complex, accommodating a whole range of different uses (cf. chapter 3). Ultimately, the present results do not license the claim that the construction's functional core has changed in any significant way. What they tell us is that certain lexical-grammatical realisations of the progressive have developed non-prototypical meanings and that they have in many cases substantially changed in frequency – nothing more and nothing less. Consequently, the observed developments should be regarded as changes in the progressive's construct-i-con that are affecting specific relations/links, but that do not alter the construction's status as a marker of progressive aspect.

7. Concluding discussion

As the final part of the present work, this chapter focuses on the central results and discusses their contribution to our conception of the progressive construction as well as to the study of language change and corpus linguists in general.

7.1. Central results

The first major insight concerns the progressive's frequency development in present-day English: In all analysed genres, the construction is more frequent today than it was in the middle of the 20th century. The only genre in which the development is not entirely clear is Face-to-face Conversation, where the combined DCPSE and FCSE data suggest a recent decline between the 1990s and today. However, this trend is not confirmed by the more comprehensive BNC and Spoken BNC2014 data, for which a continuing increase is attested in the same period. Thus, rather than having reached a saturation point, the progressive's frequency is still increasing in spoken British English.

While in the 20th century the construction used to be most frequent in Face-to-face Conversation, its level of use is now highest in media genres (Broadcast Interviews & Discussions and Spontaneous Commentary), which have traditionally been more heavily influenced by the norms of written language than everyday conversations between equals. This development suggests ongoing colloquialization (Mair 2006b: 183ff.) – a change in stylistic norms whereby frequent progressive use has firmly taken hold in more formal spoken genres. In this context, it is especially the meta-communicative uses such as *I'm not saying* that have to be mentioned since they reveal patterns of argumentative language use which cut across different spoken genres.

By compiling a new, genre-specific corpus of up-to-date spoken British English (*Freiburg Corpus of Spoken English*), which was used as a supplement to the DCPSE, and by utilising BNC and the recently released Spoken BNC2014, it became possible to reveal these and other of the progressive's developments in present-day English. This real-time, moving-window approach has resulted in a more precise description of one core aspect of contemporary English grammar.

The second major result concerns the morphosyntactic and semantic changes that have occurred alongside the frequency developments. On the one hand, it has been shown that if a frequency increase has occurred, it is connected to increased use of the progressive with the present tense and with activity/event verbs – i.e. the construction's most prototypical contexts. On the other hand, the statistical analyses have revealed that progressive use with modal auxiliaries (mainly *will*), *to*-infinitives, and especially with intermediate verbs (above all cognition/emotion/attitude and perception/sensation verbs) has also contributed to the construction's increasing use. This means that the

progressive's ongoing frequency increase is a result of both further entrenchment of the most central uses and of a slow but steady expansion of the construction's morphosyntactic and verbal paradigm. Very much in line with De Smet's (2016) theory of language change, more frequent use in conventionalised contexts increases the chance of non-prototypical/innovative uses to occur.

Crucially, the present work has not found evidence for the frequently made claim that the progressive's use with proper state verbs is increasing. The construction's level of use with existence/relationship and stance verbs shows little diachronic variation and remains on a low level in all genres. This casts doubt on the hypothesis of a progressive-to-imperfective drift, in the course of which the construction would be developing into a marker of imperfective aspect. If this was the case, we would not just expect increased usage with intermediate verbs but also with the truly stative ones. Thus, while we certainly witness a gradual extension of the progressive's verbal range, the observed changes are not dramatic enough – neither in frequency nor from a semantic point of view – to suggest a major change of the construction's functional core.

A third central finding concerns lexically specific patterns of use: By focusing on the most frequent four-word patterns and their pragmatic associations, Römer's (2005) insight that the progressive has to be regarded as a lexical-grammatical phenomenon has received further support. In addition to that, it was shown that many of the studied patterns have considerably increased in use over the past twenty years. This spread through the speech community is connected to pragmatic strengthening and increasing routinisation. Interestingly, several of the most frequent/most rapidly developing patterns are formed with *say*, serving a meta-communicative function that is ultimately based on the progressive's interpretative meaning (e.g. *I'm just saying* or *what I'm saying*). Additionally, also uses with cognition verbs (*think* and *wonder*) have become increasingly established (e.g. *I was just thinking*).

Methodologically, this work demonstrates that several aspects of a pattern's degree of routinisation can fruitfully be studied using corpus data. From frequency distributions, over the degree of compositionality, to progressive bias and verbal productivity, a whole range of phenomena can be tested that indicate how strongly the respective pattern is anchored in the speech community. While these results cannot simply be equated with a pattern's state of mental entrenchment (cf. Blumenthal-Dramé 2012), they do nonetheless relate to certain aspects of it, such as strength of representation (high token frequency) or holistic storage (low degree of compositionality). The occurrence of frozen or reduced forms of a pattern (*just saying* instead of *I'm/was just saying*) reveals a particularly advanced state of development from a constructionist point of view, indicating (completed) constructionalization.

With regard to corpus linguistics and the study of language change in general, the present study underscores the fact that spoken data is inherently very variable – probably even more than is generally assumed. While different levels of progressive use in different genres are not surprising, the results for Face-to-face Conversation are much more so: How can we observe a frequency decline between ICE-GB (1990s) and FCSE (2010s), while at the same time finding an increase between BNC (demographically sampled) and Spoken BNC2014? All of these corpora contain authentic and high-quality conversational

transcripts, qualifying as Face-to-face Conversation. While ICE-GB and FCSE are much smaller than the respective BNC counterparts, they can nonetheless be considered large enough to reliably approximate the level of use of a high-frequency construction as the progressive. If this view is accepted, the differing progressive rates can only mean that the data must vary along certain lines that are not adequately captured by the classification as Face-to-face Conversation. Apart from narrativity, whose impact has been analysed (cf. chapter 5.1.1), it seems conceivable that the level of argumentative and explanatory language use influences the progressive's occurrence rate. In this context, frequently used meta-communicative patterns such as *I'm not saying*, *are you saying* or *what I'm saying* come to mind. Since the progressive is often used to express subjective stance, the relationship between speakers might also be relevant.

Unfortunately, though, existing spoken corpus resources are not normally annotated for these variables, making analyses extremely time-consuming if not impossible. Even worse, genre distinctions are sometimes completely disregarded and quite different kinds of data are subsumed under the heading *spoken* language or *conversation*. Simply reading through the descriptions of the BYU corpora[1] – all of which are very valuable linguistic resources – illustrates the issue: The spoken data of the *Corpus of Contemporary American English* (COCA) is described as "unscripted conversation" and the data of the *Corpus of American Soap Operas* as "informal, colloquial American speech". Crucially, though, the COCA data consists of transcripts of radio and TV shows, meaning that we are actually dealing with media language. As for the soap operas, their language is entirely scripted and – as informal as it might be – does not constitute natural conversation with all its false starts, hesitations, repetitions, etc. While these remarks should by no means discourage the reader from making use of the mentioned corpus resources – after all, there are still fewer spoken corpora available than written ones –, they are simply intended to draw attention to the fact that rather diverse kinds of spoken data are sometimes described/advertised in a way that does not do justice to this variance.

7.2. Comparison to previous studies

7.2.1. Corpus-based studies

Previous studies such as Mair and Hundt (1995), Smith (2005) or Kranich (2010) analysed written data, at best supplementing it with results from spoken mini-corpora (cf. Smith 2005: 76ff.) or speech-like genres (cf. Kranich 2010: 96ff.). The present work goes beyond these studies because it is the first one analysing the progressive's diachronic development based on a large amount of transcribed spoken data.

In addition to its broader empirical base, it also breaks new ground methodologically. While the cited studies all provide reliable and detailed distributional accounts, their focus is mostly on one variable at a time and the statistical sophistication does not normally go beyond basic significance testing. In the face of the rapid development of statistical corpus-linguistic tools (cf. chapter 4.1), it was only logical to make use

[1] URL: corpus.byu.edu (visited on 12/07/2018).

of techniques such as Collostructional Analysis or Hierarchical Configural Frequency Analysis. Using them, it was possible to arrive at more fine-grained results that add to our understanding of the progressive's complex use. The application of logistic regression modelling enabled the prediction of the aspectual encoding of specific verbal patterns. Importantly, it was not just possible to quantify the influence of verb and corpus, but also to reveal a priming and recency effect of the previous aspectual encoding. In line with earlier studies by Szmrecsanyi (2006) and Gries (2005), this shows that in certain cases corpus data can indeed reflect specific psycholinguistic phenomena.

Finally, the consistent application of a cognitivist theoretical framework – usage-based Construction Grammar (CxG) – constitutes a further difference compared to previous accounts. The concept of an exemplar-based construct-i-con, in which every usage event is indexed and different levels of constructional representation can be accounted for, makes possible a meaningful interpretation of the data and of the progressive's varied usage. Since the study of frequency effects is closely connected to the usage-based CxG framework, it could fruitfully be integrated in the analysis. As hypothesised, it was found that increasing use of the progressive goes along with more frequent use of its most typical realisations (present progressive active and activity/event verbs). In line with Leech et al. (2009: 269f.), this is interpreted as a snowball effect of frequency, where increase begets further increase, making central uses more central still. The second frequency effect concerns the diversification of the construction's structural and semantic paradigm. In the wake of an increased overall token frequency, speakers' readiness to use the construction in less typical/more innovative contexts increases as well (cf. De Smet 2016), which constitutes an attraction effect of frequency. Most importantly, cognition/emotion/attitude and perception/sensation verbs have recently become more thoroughly established among the progressive's contexts of use. This is also true for modal uses with *will* and – on a low level – for combinations with *to*-infinitives. On the level of recurrent patterns of use, two further frequency effects could be attested: First, frequently used patterns tend to acquire routinised pragmatic meanings (pragmatic strengthening) that can be quite distinct from the progressive's aspectual meaning. Second, as a result of frequent use and increased routinisation, these patterns can become reduced (e.g. *I'm/was just saying –> just saying*), which constitutes another well-known frequency effect (cf. chapter 2.1.2).

7.2.2. Reference grammars and usage guides

Treatments of the progressive in reference grammars and usage guides have not so much focused on quantitative aspects of the construction's distribution but have mainly provided very detailed semantic accounts. Above all, there is Leech (2004), who devotes a whole chapter to the progressive's use in present-day English. Additionally, there are frequently cited accounts such as the ones by Quirk et al. (1985) or Huddleston and Pullum (2002). What these treatments have in common is that they acknowledge the existence of a range of functions that the progressive can express in addition to its aspectual meaning. For example, both Leech (2004: 31f.) and Quirk et al. (1985: 210) discuss non-aspectual past progressive uses that qualify as instances of the recentness progres-

sive. Similarly, Huddleston and Pullum (2002: 105) and Leech (2004: 22) make reference to the interpretative progressive. The matter-of-course future (*will* plus progressive) is also addressed in each of the mentioned publications.

Especially Leech (2004: 23ff.) makes a very detailed effort to capture the construction's actual verbal range, clearly showing in which contexts supposedly anti-progressive verbs (e.g. perception or cognition verbs) can be used naturally. Even though he does not provide quantitative evidence, he states that "[i]t seems as if usage in this area is not always logical and systematic, because the language itself is gradually extending the use of the Progressive" (2004: 25).

The results of the present study corroborate these analyses by providing a detailed picture of how exactly certain of these features of progressive use (e.g. co-occurrence with perception/sensation or cognition verbs) have taken hold in the speech community in the course of the past few decades. Thus, they lend quantitative support to previously made qualitative observations.

As a corpus-based grammar, the Longman Grammar (Biber et al. 1999) is more descriptive and less analytical in nature than for example Quirk et al. (1985) or Huddleston and Pullum (2002). As part of its chapter on frequently used lexical expressions (1999: 987ff.), several lexically specific progressive patterns are identified, hinting at the construction's lexical-grammatical nature (e.g. *that's what I'm saying* or *what are you talking about*). While the present work has found similar patterns, it has moved beyond mere identification by providing detailed discourse-pragmatic analyses combined with further quantitative assessment, including the patterns' recent diachronic development in spoken conversation.

The findings of this study will not necessitate a complete re-writing of existing grammars and usage-guides. However, this work has presented a detailed analysis of the construction's current distributional and functional status in spoken British English, covering both the most central and also special aspects of use. Quantitative data has been provided for all phenomena under study, thereby offering a well-grounded up-to-date account that goes beyond the often intuitive descriptions in existing works of reference.

7.3. Application and suggestions for further research

The most obvious and important application of this work's results is the teaching of English as a second language. Many learners struggle to come to grips with the progressive's complex functionality and verbal range. While clearly aspectual uses with dynamic verbs do not seem to be the major problem, it is the less typical, not clearly aspectual uses that are most challenging. As we have seen, however, some of these uses are gaining ground, such as the matter-of-course future, the recentness progressive or uses with intermediate verbs. It is thus imperative that teaching resources and pedagogical grammars be updated accordingly, doing justice to the changing use of the construction. Furthermore, many teachers still seem to adhere to the frequently made claim that verbs such as *think* or *feel* must not normally be used with the progressive.

Instead, teaching materials should highlight the fact that this is too general a statement and should provide examples of the right contexts of use. Frequent progressive patterns and their discourse-pragmatic meanings should also be part of pedagogical grammars of English, acknowledging the insight that grammar and lexis are interdependent and equipping pupils with an authentic conception of the progressive.

With regard to further research, it is the psycholinguistic perspective that should be brought into focus. For example, it would be interesting to compare typical progressive uses such as *He is currently working* to uses such as *I'll be meeting him on Monday* (matter-of-course future) or *Tim was just saying [...]* (recentness progressive). How do speakers judge these uses regarding their aspectual meaning? Is there still an underlying sense of progressivity that unites them or is the perceived similarity simply a formal one? Likewise, it should be tested whether frequent patterns of use such as *I was just thinking* have acquired construction status – i.e. are stored holistically – or whether they are simply processed faster than less frequent realisations of the progressive. Studies such as Caldwell-Harris et al. (2012) – using a perceptual identification task – or Schmitt et al. (2004) – using a dictation test/oral response task – have shown that experimental paradigms exist by means of which chunk status can indeed be operationalised.

Last but not least, that most dangerous step in studies of diachronic change, namely predicting the future, does not seem to be entirely beyond our reach. Given the vast amount of long-term and short-term diachronic data, probabilistic modelling of possible future developments might be a realistic alternative to waiting for two decades to carry out the next "moving window" study.

A. Appendix

A.4. Corpus texts *Freiburg Corpus of Spoken English* (FCSE)

Genre	Text-ID	Data	Year	Speakers	Words
FFC	GRK_FFC_1	*Listening Project*	2014	2	4254
	GRK_FFC_2	*Listening Project*	2014	2	13603
	GRK_FFC_3	*Listening Project*	2014	2	9507
	GRK_FFC_4	*Listening Project*	2012	2	8060
	GRK_FFC_5	*Listening Project*	2014	2	8426
	GRK_FFC_6	*Listening Project*	2014	2	5934
	GRK_FFC_7	*Listening Project*	2014	2	6865
	GRK_FFC_8	*Listening Project*	2013	2	6182
	GRK_FFC_9	*Listening Project*	2012	2	7999
	GRK_FFC_10	*Listening Project*	2014	2	10142
	GRK_FFC_11	*Listening Project*	2013	2	8062
	GRK_FFC_12	*Listening Project*	2012	2	8334
	GRK_FFC_13	*Listening Project*	2013	2	6381
	GRK_FFC_14	*Listening Project*	2012	2	7058
	GRK_FFC_15	*Listening Project*	2014	2	6274
	GRK_FFC_16	*Listening Project*	2013	2	7408
	GRK_FFC_17	*Listening Project*	2012	2	7949
	GRK_FFC_18	*Listening Project*	2012	2	9379
	GRK_FFC_19	*Listening Project*	2013	2	7873
					149,690
BD	GRK_BD_1	BBC *Ouch*	2012	5	5531
	GRK_BD_2	BBC *Ouch*	2012	6	6021
	GRK_BD_3	BBC *Ouch*	2012	6	5544
	GRK_BD_4	BBC *Ouch*	2014	6	5592
	GRK_BD_5	BBC *Ouch*	2013	4	5019
	GRK_BD_6	BBC *Ouch*	2013	5	5720
	GRK_BD_7	BBC *Ouch*	2014	6	5470
	GRK_BD_8	BBC *Ouch*	2015	6	6054
	GRK_BD_9	BBC *Ouch*	2016	5	5243
	GRK_BD_10	BBC *Ouch*	2016	5	5736
					55,930
BI	GRK_BI_1	BBC *In Touch*	2015	3	1341
	GRK_BI_2_a	BBC *In Touch*	2015	2	487

	GRK_BI_2_b	BBC *In Touch*	2015	4	860
	GRK_BI_3	BBC *In Touch*	2015	2	2461
	GRK_BI_4_a	BBC *In Touch*	2015	2	1225
	GRK_BI_4_b	BBC *In Touch*	2015	2	463
	GRK_BI_4_c	BBC *In Touch*	2015	2	1031
	GRK_BI_5_a	BBC *In Touch*	2015	2	422
	GRK_BI_5_b	BBC *In Touch*	2015	2	330
	GRK_BI_5_c	BBC *In Touch*	2015	2	1283
	GRK_BI_6_a	BBC *In Touch*	2015	2	1180
	GRK_BI_6_b	BBC *In Touch*	2015	2	1604
	GRK_BI_7	BBC *In Touch*	2015	3	3096
	GRK_BI_8	BBC *Andrew Marr Show**	2015	2	1839
	GRK_BI_9	BBC *Andrew Marr Show**	2015	2	1868
	GRK_BI_10	BBC *Andrew Marr Show**	2015	2	1306
	GRK_BI_11	BBC *Andrew Marr Show**	2015	2	2199
	GRK_BI_12	BBC *Andrew Marr Show**	2013	2	2782
	GRK_BI_13	BBC *Andrew Marr Show**	2013	2	1361
	GRK_BI_14	BBC *Andrew Marr Show**	2013	2	815
	GRK_BI_15	BBC *Andrew Marr Show**	2013	2	2639
	GRK_BI_16	BBC *Andrew Marr Show**	2013	2	4086
	GRK_BI_17	BBC *Andrew Marr Show**	2013	2	1167
	GRK_BI_18	BBC *Andrew Marr Show*	2015	2	1868
	GRK_BI_19	BBC *Andrew Marr Show*	2015	2	2315
	GRK_BI_20	BBC *Andrew Marr Show*	2015	2	2035
	GRK_BI_21	BBC *Andrew Marr Show*	2015	2	1697
	GRK_BI_22	BBC *Andrew Marr Show*	2016	2	2086
	GRK_BI_23	BBC *Andrew Marr Show*	2016	2	1813
	GRK_BI_24	BBC *Andrew Marr Show*	2016	2	2447
	GRK_BI_25	BBC *Andrew Marr Show*	2016	2	1899
					52,005
PL	GRK_PL_1	Hansard House of Commons	2015	13	4989
	GRK_PL_2	Hansard House of Commons	2015	19	5030
	GRK_PL_3	Hansard House of Commons	2015	19	4718
	GRK_PL_4	Hansard House of Commons	2015	36	6712
	GRK_PL_5	Hansard House of Commons	2015	13	5341
	GRK_PL_6	Hansard House of Commons	2015	19	4304
	GRK_PL_7	Hansard House of Commons	2015	21	4820
	GRK_PL_8	Hansard House of Commons	2015	25	6364
	GRK_PL_9	Hansard House of Commons	2015	17	4882
	GRK_PL_10	Hansard House of Commons	2015	20	4891
					52,051
PS	GRK_PS_1	Political Speech	2014	1	5123
	GRK_PS_2	Political Speech	2014	1	5207
	GRK_PS_3	Political Speech	2014	1	5358

	GRK_PS_4	Political Speech	2015	1	5143
	GRK_PS_5	Political Speech	2015	1	5506
	GRK_PS_6	Political Speech	2015	1	4169
	GRK_PS_7	Political Speech	2016	1	5681
	GRK_PS_8	Public Speech	2014	1	1205
	GRK_PS_9	Sermon	2014	1	1544
	GRK_PS_10	Sermon	2013	1	1534
	GRK_PS_11	Sermon	2014	1	1515
	GRK_PS_12	Sermon	2014	1	1442
	GRK_PS_13	Public Lecture	2014	1	5007
	GRK_PS_14	Public Lecture	2014	1	4110
	GRK_PS_15	Public Lecture	2014	1	3923
	GRK_PS_16	Public Lecture	2014	1	5070
	GRK_PS_17	Public Lecture	2014	2	5250
	GRK_PS_18	Public Lecture	2015	1	5369
	GRK_PS_19	Public Lecture	2015	1	5031
	GRK_PS_20	Public Lecture	2015	1	4949
	GRK_PS_21	Public Lecture	2015	1	5033
	GRK_PS_22	Public Lecture	2015	1	5113
	GRK_PS_23	Public Lecture	2013	1	5077
	GRK_PS_24	Public Lecture	2013	1	5403
					102,762
SC	GRK_SC_1	Sports Commentary	2015	2	4490
	GRK_SC_2	Sports Commentary	2014	2	4855
	GRK_SC_3	Sports Commentary	2014	2	4961
	GRK_SC_4	Sports Commentary	2015	2	4796
	GRK_SC_5	Sports Commentary	2015	2	4623
	GRK_SC_6	Sports Commentary	2016	6	4851
	GRK_SC_7	Ceremonial Commentary	2016	4	4771
	GRK_SC_8	Ceremonial Commentary	2015	3	4695
	GRK_SC_9	Ceremonial Commentary	2014	1	4567
	GRK_SC_10	Sports Commentary	2016	5	4931
					47,540
					459,978

Table A.1.: Texts *Freiburg Corpus of Spoken English* (FCSE). FFC = Informal Face-to-face Conversation, BD = Broadcast Discussions, BI = Broadcast Interviews, PL = Parliamentary Language, PS = Prepared Speech, SC = Spontaneous Commentary (***** Interviews conducted by different stand-in moderators).

A.5. The progressive's frequency development

	LLC			ICE-GB			FCSE			Difference LLC - ICE-GB		Difference ICE-GB - FCSE		Difference LLC - FCSE	
Genre	Raw freq	word no.	**freq htw**	Raw freq	word no.	**freq htw**	Raw freq	word no.	**freq htw**	sign. level	change in %	sign. level	change in %	sign. level	**change in %**
FFC	1384	218307	**634**	1523	185537	**821**	1031	149690	**689**	***	+29.5%	***	–16.1%	*	**+8.7%**
BI	126	20899	**603**	121	22147	**546**	503	52005	**967**	*ns.*	–9.5%	***	+77.1%	***	**+60.4%**
BD	288	45237	**637**	347	43920	**790**	499	55930	**892**	*	+24.0%	*ns.*	+12.9%	***	**+40.0%**
PL	43	10494	**410**	42	10589	**397**	468	52051	**899**	*ns.*	–3.2%	***	+126.4%	***	**+119.3%**
PS	128	31182	**410**	153	32393	**472**	506	102762	**492**	*ns.*	+15.1%	*ns.*	+4.2%	*ns.*	**+20.0%**
SC	224	46426	**482**	293	48955	**599**	406	47540	**854**	*	+24.3%	***	+42.6%	***	**+77.2%**

Table A.2.: The progressive's frequency development in DCPSE and FCSE (Chi-square test for statistical significance: *** p<0.001, ** p<0.01, * p<0.05) htw = hundred thousand words, FFC = Informal Face-to-face Conversation, BI = Broadcast Interviews, BD = Broadcast Discussions, PL = Parliamentary Language, PS = Prepared Speech, SC = Spontaneous Commentary.

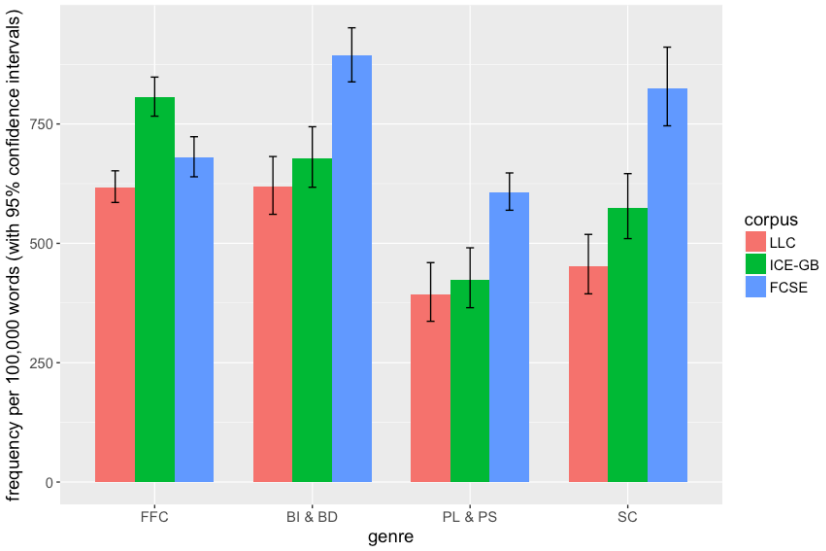

Figure A.1.: Progressive frequency by corpus and genre (active progressives only). FFC = Informal Face-to-face Conversation, BI & BD = Broadcast Interviews and Discussions, PL & PS = Parliamentary Language and Prepared Speech, SC = Spontaneous Commentary.

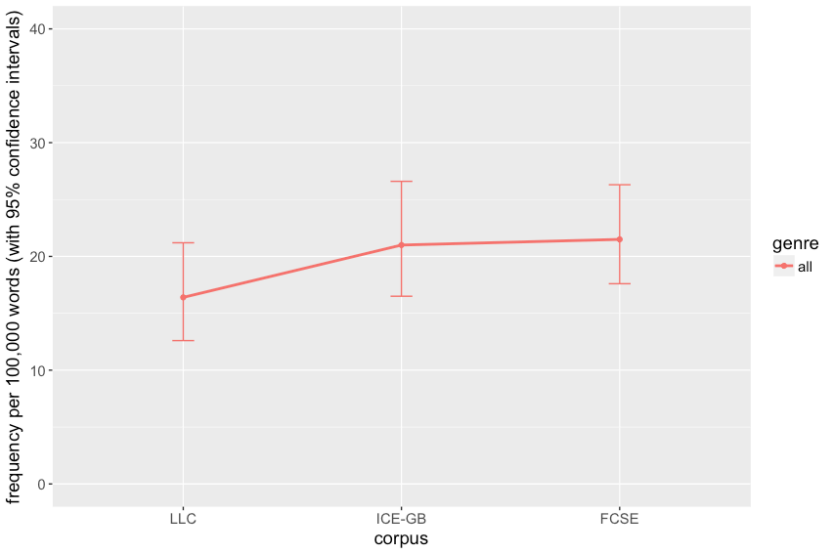

Figure A.2.: Development of the progressive passive (no genre distinction).

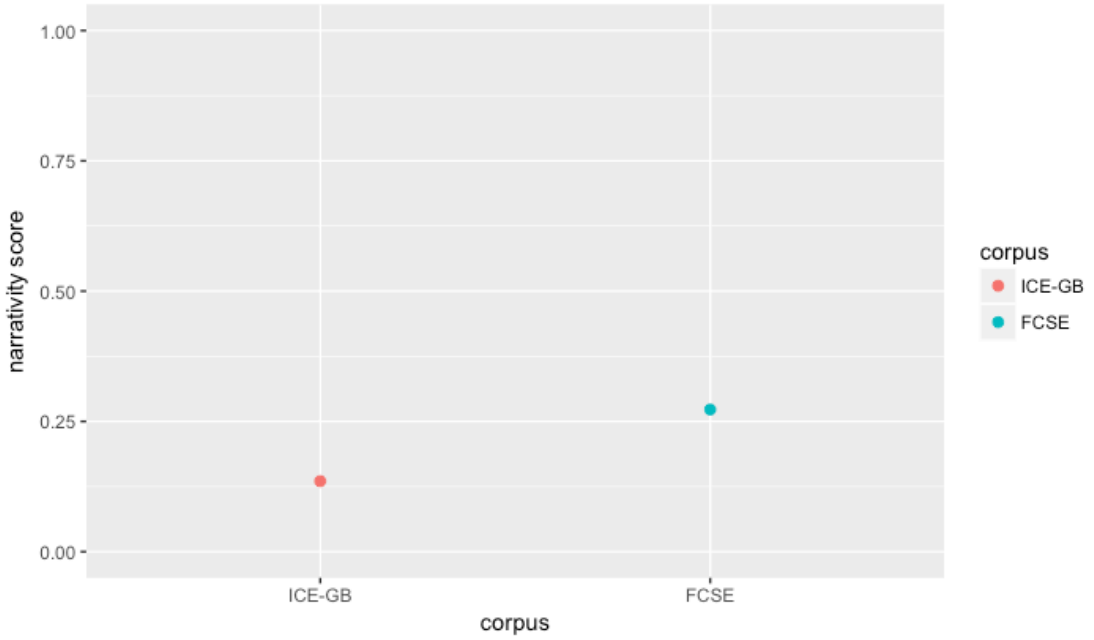

Figure A.3.: Face-to-face Conversation ICE-GB & FCSE: Narrativity scores (based on the word count of narrative passages in the analysed corpus texts. Maximum = 1 (100% of all words), minimum = 0 (0% of all words)).

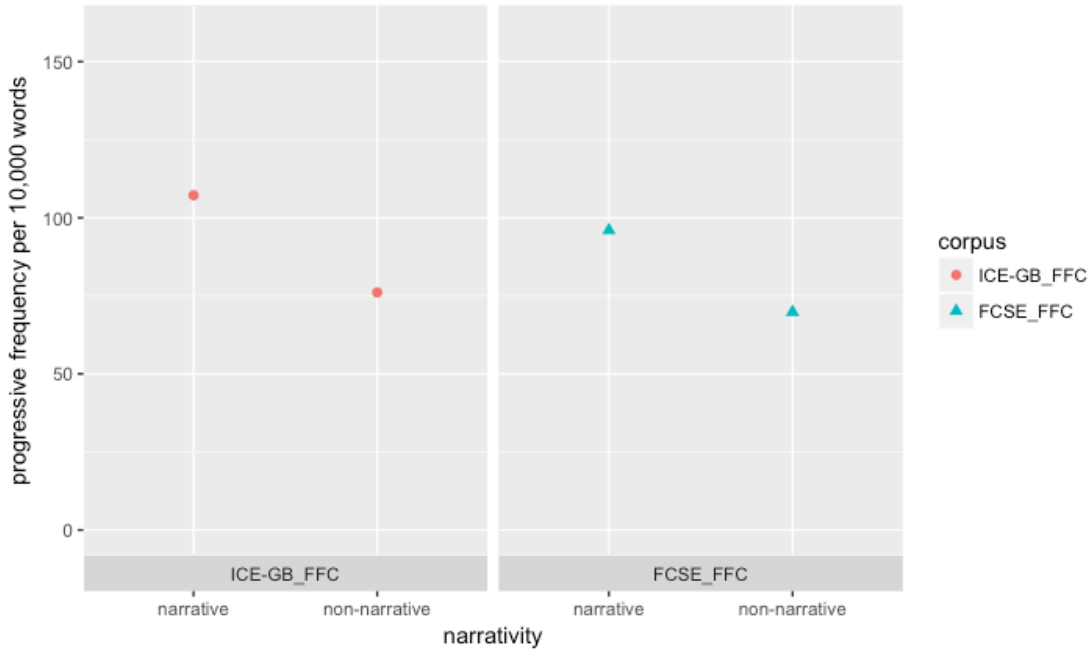

Figure A.4.: ICE-GB & FCSE Face-to-face Conversation: Progressive frequency in narrative and non-narrative passages.

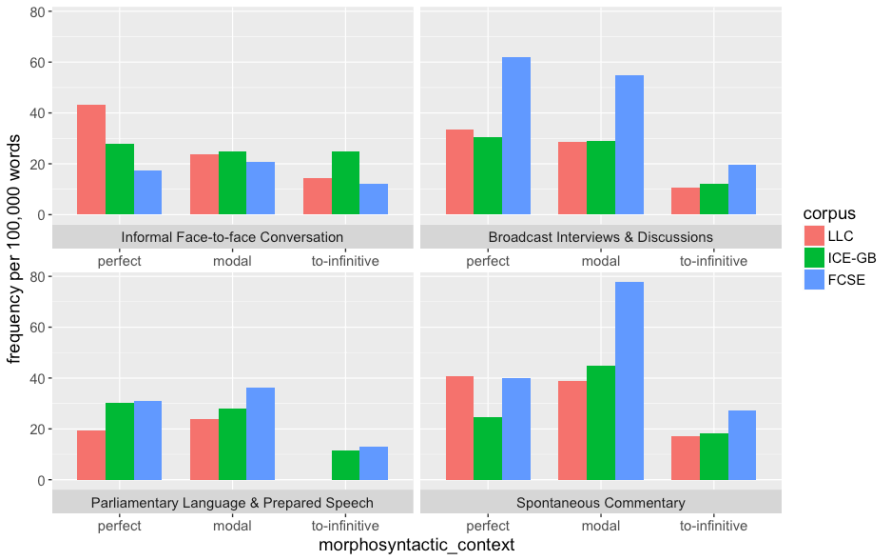

Figure A.5.: Morphosyntactic context by corpus and genre (without present and past progressive) (active progressives only). per = present & past perfect, mo = modal, toI = *to*-Infinitive.

genre	morphosyntactic context	raw freq LLC	freq LLC htw	raw freq ICE	freq ICE htw	raw freq FCSE	freq FCSE htw	Difference LLC/ICE sign. level	% change LLC/ICE	Difference ICE/FCSE sign. level	% change ICE/FCSE	Difference LLC/FCSE sign. level	% change LLC/FCSE
FFC	present	715	328	983	530	593	396	***	+61.8%	***	–25.2%	***	+21%
	past	457	209	369	199	350	234	*ns.*	–5%	*	+17.5%	*ns.*	+11.7%
	perfect	94	43	52	28	26	17	*	–35%	*ns.*	–37.9%	***	–59.6%
	modal	52	24	46	25	31	21	*ns.*	+4.2%	*ns.*	–16.5%	*ns.*	–13%
	to-inf	31	14	46	25	18	12	*	+74.6%	**	–51.6%	*ns.*	–15.5%
BI&BD	present	272	411	273	413	675	625	*ns.*	+0.5%	***	+51.4%	***	+52.1%
	past	89	135	128	194	142	132	**	+43.9%	**	–32.1%	*ns.*	–2.2%
	perfect	22	33	20	30	67	62	*ns.*	–9%	**	+105%	*	+86.5%
	modal	19	29	19	29	59	55	*ns.*	+0.3%	*	+89.9%	*	+90.6%
	to-inf	7	11	8	12	21	20	*ns.*	+14.2%	*ns.*	+61.2%	*ns.*	+84%
PL&PS	present	76	182	114	265	658	425	*	+45.4%	***	+60.3%	***	+133%
	past	70	168	38	88	158	102	**	–47.4%	*ns.*	+15.5%	***	–39.2%
	perfect	8	19	13	30	48	31	*ns.*	+57.3%	*ns.*	+2.6%	*ns.*	+61.5%
	modal	10	24	12	28	56	36	*ns.*	+16.2%	*ns.*	+29.7%	*ns.*	+50.8%
	to-inf	0	0	5	12	20	13	*ns.*	Inf	*ns.*	+11.2%	*	Inf
SC	present	155	334	209	427	263	553	*	+27.9%	**	+29.6%	***	+65.7%
	past	10	22	29	59	60	126	**	+175.3%	***	+113.2%	***	+487%
	perfect	19	41	12	24	19	40	*ns.*	–40.1%	*ns.*	+63.3%	*ns.*	–2.2%
	modal	18	39	22	45	37	78	*ns.*	+15.7%	*	+73.3%	*	+100.5%
	to-inf	8	17	9	18	13	27	*ns.*	+7%	*ns.*	+48.4%	*ns.*	+58.7%

Table A.3.: The progressives's morphosyntactic context (active progressives only): Development in DCPSE and FCSE (Fisher's exact test: *** p<0.001, ** p<0.01, * p<0.05) ICE = ICE-GB, htw = hundred thousand words, to-inf = to-infinitive, FFC = Informal Face-to-face Conversation, BI&BD = Broadcast Interviews & Discussions, PL&PS = Parliamentary Language & Prepared Speech, SC = Spontaneous Commentary; frequencies per htw have been rounded after calculation of percental change.

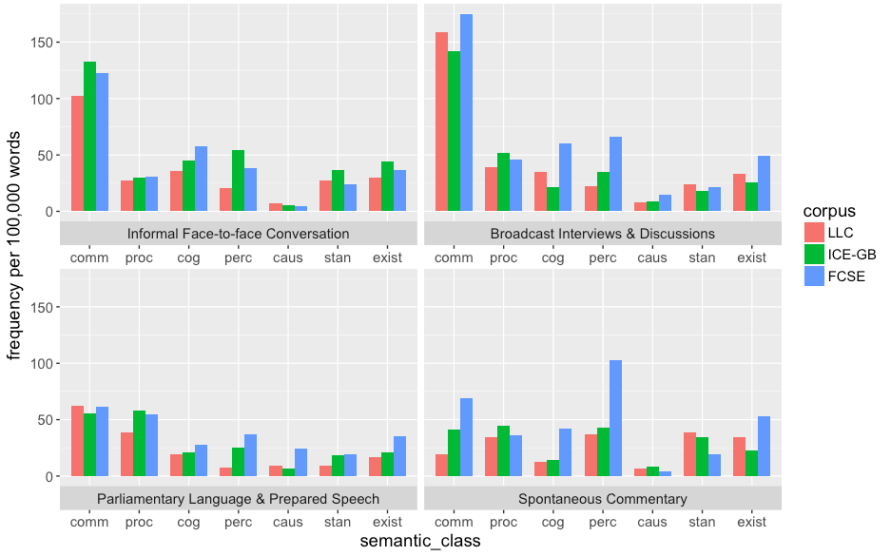

Figure A.6.: Verb class by corpus and genre (without activity/event verbs) (active progressives only). caus = causative, cog = cognition/emotion/attitude, comm = communication, exist = existence/relationship, perc = perception/sensation, proc = process/aspectual, stan = stance.

genre	semantic class	raw freq LLC	freq LLC htw	raw freq ICE	freq ICE htw	raw freq FCSE	freq FCSE htw	Difference LLC/ICE sign. level	% change LLC/ICE	Difference ICE/FCSE sign. level	% change ICE/FCSE	Difference LLC/FCSE sign. level	% change LLC/FCSE
FFC	act	804	368	848	457	546	365	***	+24.1%	***	–20.2%	*ns.*	–1%
	comm	223	102	247	133	184	123	**	+30.4%	*ns.*	–7.7%	*ns.*	+20.4%
	cog	79	36	84	45	87	58	*ns.*	+25.1%	*ns.*	+28.3%	**	+60.5%
	exist	65	30	82	44	55	37	*	+48.3%	*ns.*	–17%	*ns.*	+23.2%
	proc	60	28	56	30	46	31	*ns.*	+9.8%	*ns.*	+1.7%	*ns.*	+11.6%
	stan	59	27	68	37	36	24	*ns.*	+35.9%	*	–34.6%	*ns.*	–11.1%
	perc	44	20	101	54	57	38	***	+169.3%	*	–30%	**	+88.6%
	caus	15	7	10	5	7	5	*ns.*	–21.7%	*ns.*	–13%	*ns.*	–31.9%
BI&BD	act	197	298	248	375	497	460	*	+26%	**	+22.7%	***	+54.6%
	comm	105	159	94	142	189	175	*ns.*	–10.4%	*ns.*	+23%	*ns.*	+10.3%
	cog	23	35	14	21	65	60	*ns.*	–39.1%	***	+184%	*	+73%
	exist	22	33	17	26	53	49	*ns.*	–22.8%	*	+91.1%	*ns.*	+47.4%
	proc	26	39	34	52	50	46	*ns.*	+31%	*ns.*	–10.1%	*ns.*	+17.8%
	stan	16	24	12	18	23	21	*ns.*	–24.8%	*ns.*	+17%	*ns.*	–12%
	perc	15	23	23	35	71	66	*ns.*	+53.3%	**	+89.1%	***	+189.9%
	caus	5	8	6	9	16	15	*ns.*	+19.7%	*ns.*	+62.6%	*ns.*	+94.7%
PL&PS	act	96	230	93	216	539	348	*ns.*	–6%	***	+60.9%	***	+51.2%
	comm	26	62	24	56	95	61	*ns.*	–10.6%	*ns.*	+10%	*ns.*	–1.6%
	cog	8	19	9	21	43	28	*ns.*	+8.9%	*ns.*	+33%	*ns.*	+44.8%
	exist	7	17	9	21	54	35	*ns.*	+24.4%	*ns.*	+67%	*ns.*	+107.7%
	proc	16	38	25	58	84	54	*ns.*	+51.6%	*ns.*	–6.7%	*ns.*	+41.4%
	stan	4	10	8	19	30	19	*ns.*	+93.8%	*ns.*	+4.3%	*ns.*	+102.1%
	perc	3	7	11	26	57	37	*ns.*	+255.6%	*ns.*	+43.7%	***	+411.1%
	caus	4	10	3	7	38	24	*ns.*	–27.1%	*	+250%	*ns.*	+155.2%
SC	act	125	269	179	366	237	498	**	+35.8%	**	+36.4%	***	+85.2%
	comm	9	19	20	41	33	69	*ns.*	+110.8%	*ns.*	+69.7%	***	+257.7%
	cog	6	13	7	14	20	42	*ns.*	+10.9%	*	+194.4%	**	+226.4%

exist	16	34	11	22	25	53	*ns.*	–34.8%	*	+133.8%	*ns.*	+52.5%
proc	16	34	22	45	17	36	*ns.*	+30.1%	*ns.*	–20.3%	*ns.*	+3.8%
stan	18	39	17	35	9	19	*ns.*	–10.6%	*ns.*	–45.5%	*ns.*	–51.3%
perc	17	37	21	43	49	103	*ns.*	+17.2%	***	+140.3%	***	+181.7%
caus	3	6	4	8	2	4	*ns.*	+26.2%	*ns.*	–48.8%	*ns.*	–35.4%

Table A.4.: The progressives's verbal paradigm (active progressives only): Development in DCPSE and FCSE (Fisher's exact test: *** p<0.001, ** p<0.01, * p<0.05) ICE = ICE-GB, htw = hundred thousand words, act = activity/event, comm = communication, exist = existence/relationship, proc = process/aspectual, stan = stance, perc = perception/sensation, caus = causative, FFC = Informal Face-to-face Conversation, BI&BD = Broadcast Interviews & Discussions, PL&PS = Parliamentary Language & Prepared Speech, SC = Spontaneous Commentary; frequencies per htw have been rounded after calculation of percental change.

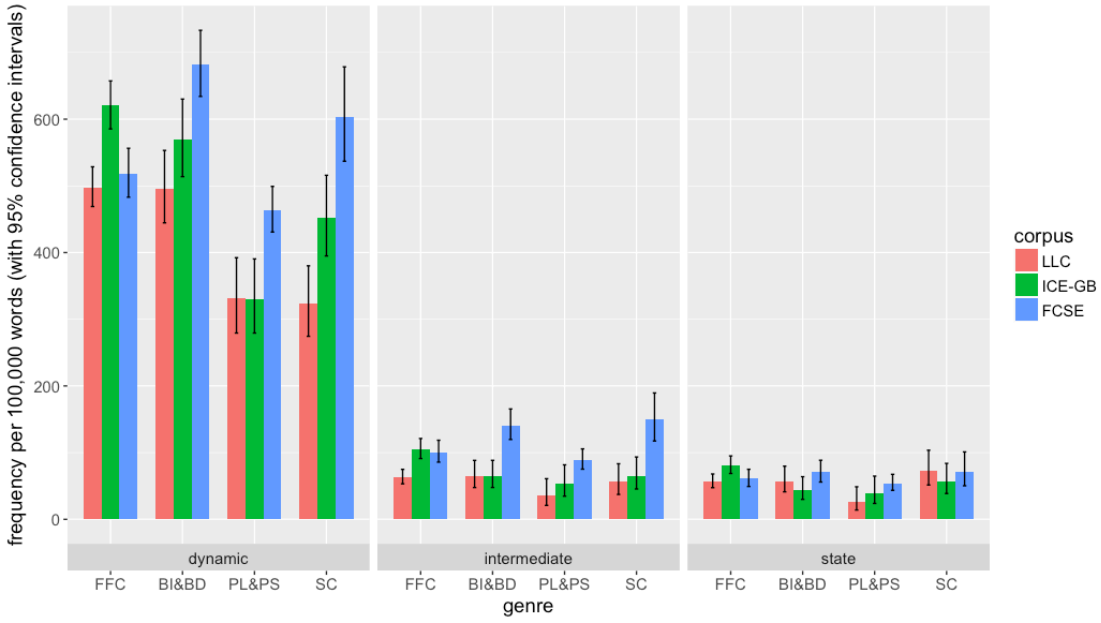

Figure A.7.: Verb class by corpus and genre (active progressives only). dynamic = activity/event + process/aspectual + communication, intermediate = causative + cognition/emotion/attitude + perception/sensation, state = existence/relationship + stance; FFC = Informal Face-to-face Conversation, BI&BD = Broadcast Interviews & Discussions, PL&PS = Parliamentary Language & Prepared Speech, SC = Spontaneous Commentary.

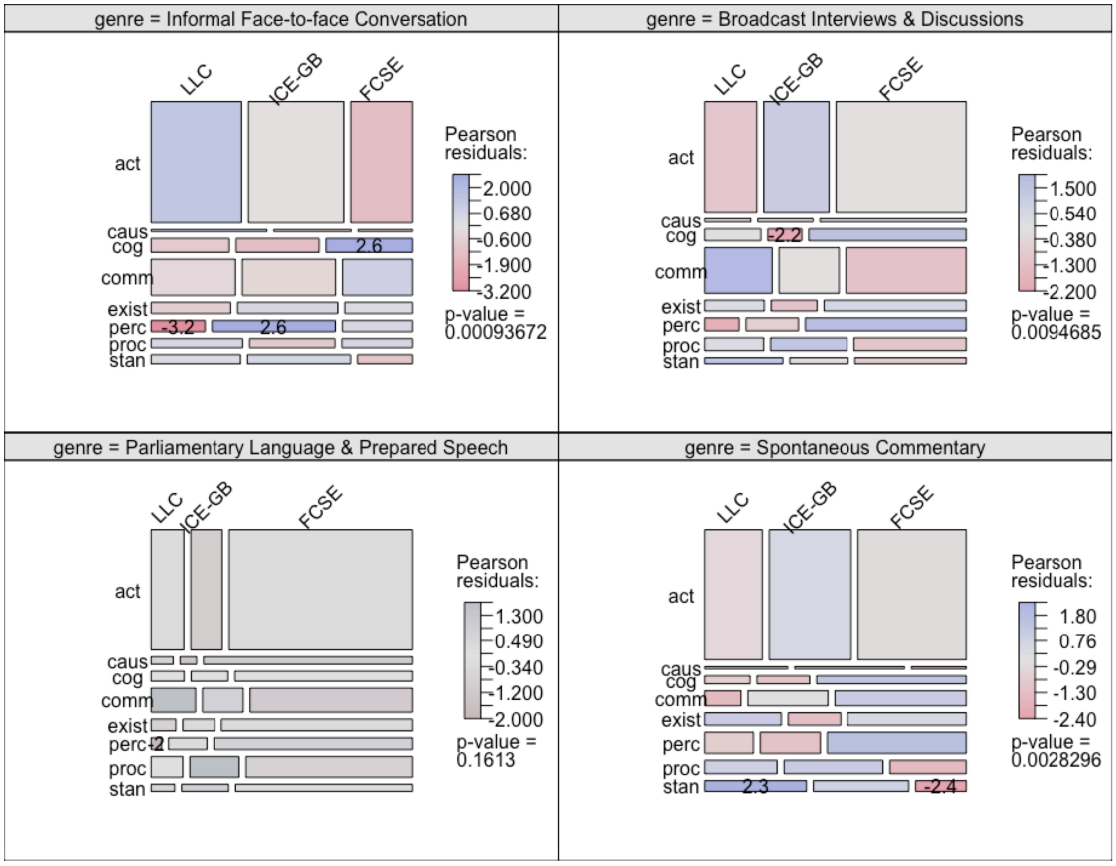

Figure A.8.: Verb class by corpus and genre (active progressives only). act = activity/event, caus = causative, cog = cognition/emotion/attitude, comm = communication, exist = existence/relationship, perc = perception/sensation, proc = process/aspectual, stan = stance. Numbers and shading indicate Pearson Residuals.

A.6. Frequent patterns of use

Realisation	freq BNC94	freq BNC94/pmw	freq BNC14	freq BNC14/pmw	% change BNC94/14	sign. level
I'm not going	115	22.9	170	14.9	–34.9%	***
I'm not doing	71	14.2	155	13.6	–4.2%	*ns.*
I'm not saying	70	14.0	232	20.3	+45.0%	**
I'm not having	51	10.2	74	6.5	–36.3%	*
I'm not getting	30	6.0	52	4.6	–23.3%	*ns.*
I'm not paying	28	5.6	42	3.7	–33.9%	*ns.*
I'm not being	27	5.4	104	9.1	+68.5%	*
I'm not telling	25	5.0	29	2.5	–50.0%	*
I'm not coming	23	4.6	22	1.9	–58.7%	**
I'm not looking	20	4.0	40	3.5	–12.5%	*ns.*
I'm not buying	15	3.0	17	1.5	–50.0%	*ns.*
I'm not joking	15	3.0	19	1.7	–43.3%	*ns.*
I'm not giving	13	2.6	22	1.9	–26.9%	*ns.*
I'm not playing	13	2.6	9	0.8	–69.2%	**
I'm not talking	13	2.6	45	3.9	+50.0%	*ns.*
I'm not eating	11	2.2	33	2.9	+31.8%	*ns.*
I'm not kidding	11	2.2	10	0.9	–59.1%	*ns.*
I'm not putting	11	2.2	21	1.8	–18.2%	*ns.*
I'm not taking	11	2.2	36	3.2	+45.5%	*ns.*
I'm not leaving	9	1.8	7	0.6	–66.7%	*

Table A.5.: Frequency development of the *I am/'m not -ing* pattern (20 most frequent realisations) (Fisher's exact test: *** p<0.001, ** p<0.01, * p<0.05) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

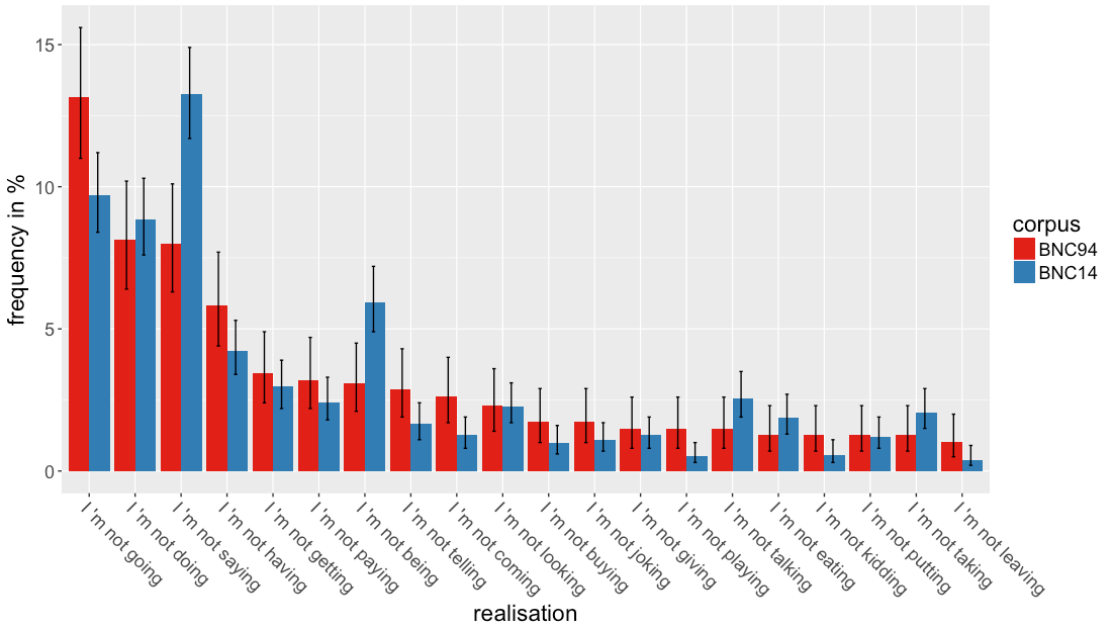

Figure A.9.: Frequency profile of the *I am/'m not -ing* pattern (20 most frequent realisations) (freq. in %; with 95% confidence intervals) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

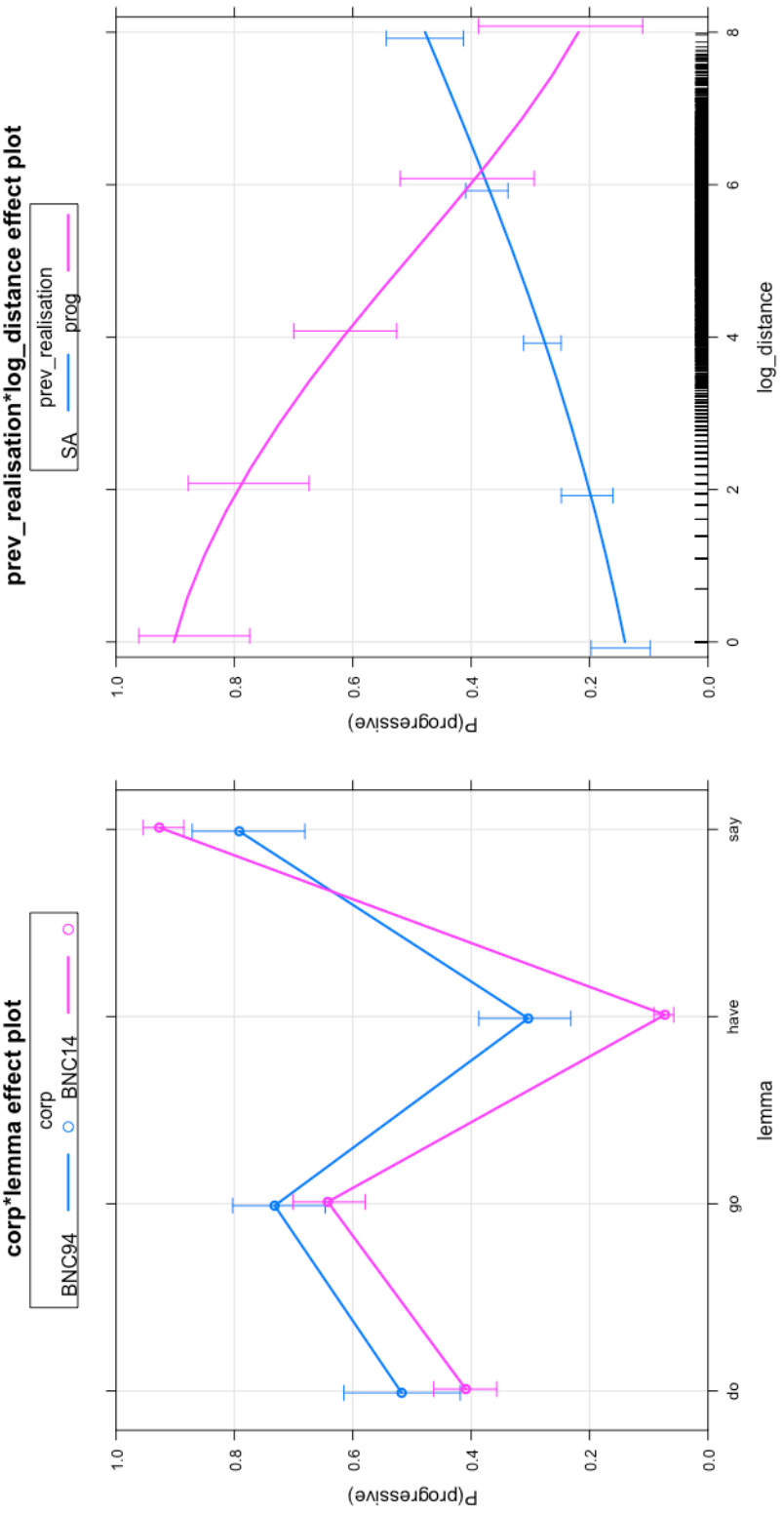

Figure A.10.: Binary logistic regression *I am/'m not -ing*: Aspectual choice – all effects.

Realisation	freq BNC94	freq BNC94/pmw	freq BNC14	freq BNC14/pmw	% change BNC94/14	sign. level
I'm just trying	36	7.2	91	8.0	+11.1%	*ns*
I was just thinking	35	7.0	138	12.1	+72.9%	**
I'm just saying	33	6.6	197	17.2	+160.6%	***
I'm just going	23	4.6	56	4.9	+6.5%	*ns*
I was just saying	22	4.4	68	6.0	+36.4%	*ns*
I'm just wondering	20	4.0	54	4.7	+17.5%	*ns*
I'm just thinking	17	3.4	91	8.0	+135.3%	***
I'm just looking	16	3.2	34	3.0	–6.3%	*ns*
I was just wondering	13	2.6	39	3.4	+30.8%	*ns*
I'm just getting	11	2.2	29	2.5	+13.6%	*ns*
I was just looking	11	2.2	36	3.2	+45.5%	*ns*
I'm just doing	10	2.0	22	1.9	–5.0%	*ns*
I'm just having	10	2.0	22	1.9	–5.0%	*ns*
I was just getting	8	1.6	11	1.0	–37.5%	*ns*
I was just talking	8	1.6	17	1.5	–6.3%	*ns*
I'm just making	7	1.4	13	1.1	–21.4%	*ns*
I'm just waiting	7	1.4	15	1.3	–7.1%	*ns*
I'm just asking	6	1.2	7	0.6	–50.0%	*ns*
I'm just putting	6	1.2	8	0.7	–41.7%	*ns*
I'm just seeing	6	1.2	7	0.6	–50.0%	*ns*

Table A.6.: Frequency development of the *I BE just -ing* pattern (20 most frequent realisations) (Fisher's exact test: *** p<0.001, ** p<0.01, * p<0.05) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

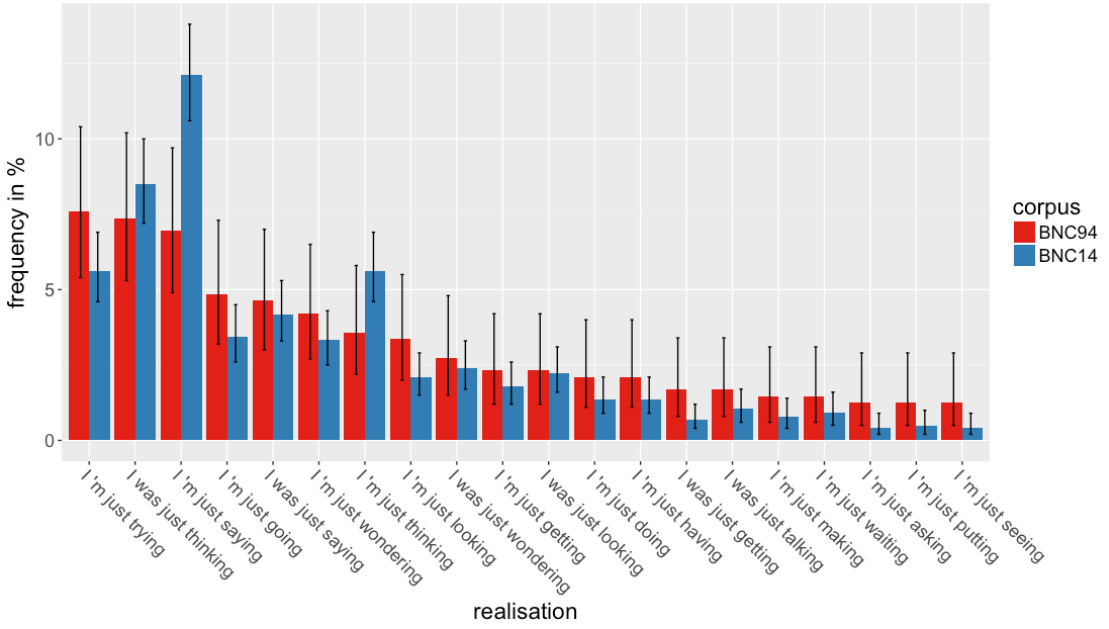

Figure A.11.: Frequency profile of the *I BE just -ing* pattern (20 most freq. realisations) (freq in %; with 95% confidence intervals) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

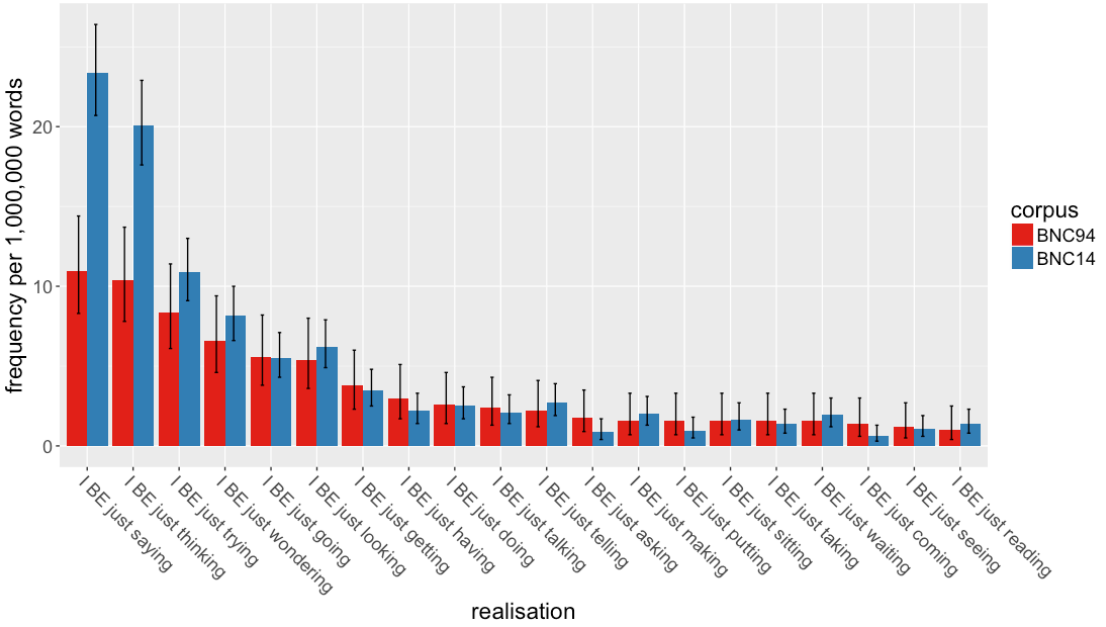

Figure A.12.: Frequency profile of the *I BE just -ing* pattern (20 most freq. realisations with 95% confidence intervals, no tense distinction) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

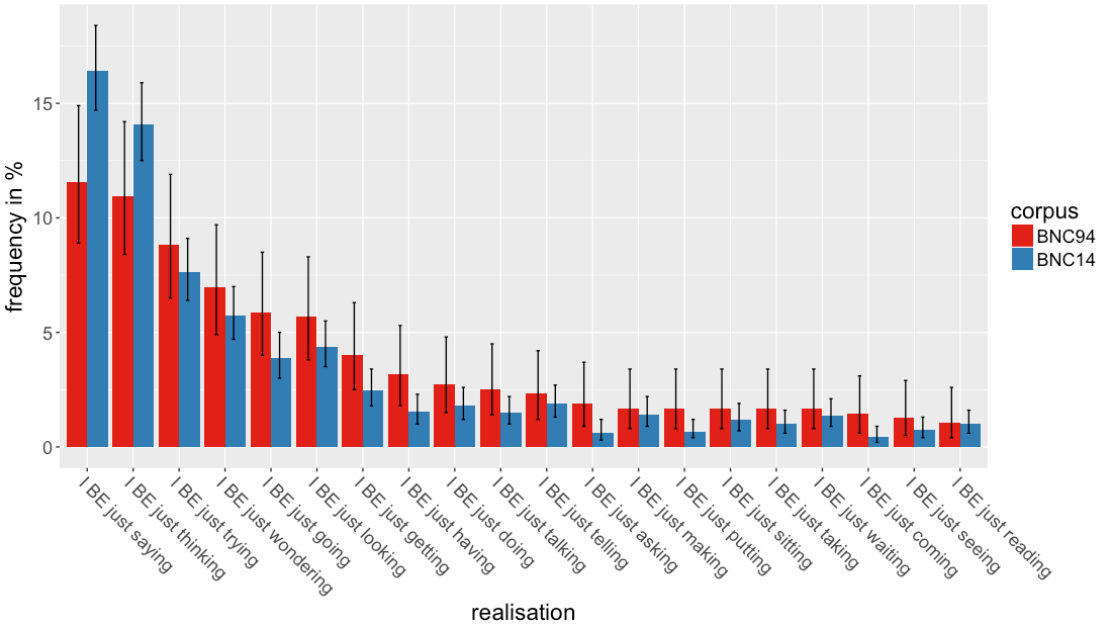

Figure A.13.: Frequency profile of the *I BE just -ing* pattern (20 most freq. realisations, no tense distinction) (freq in %; with 95% confidence intervals) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

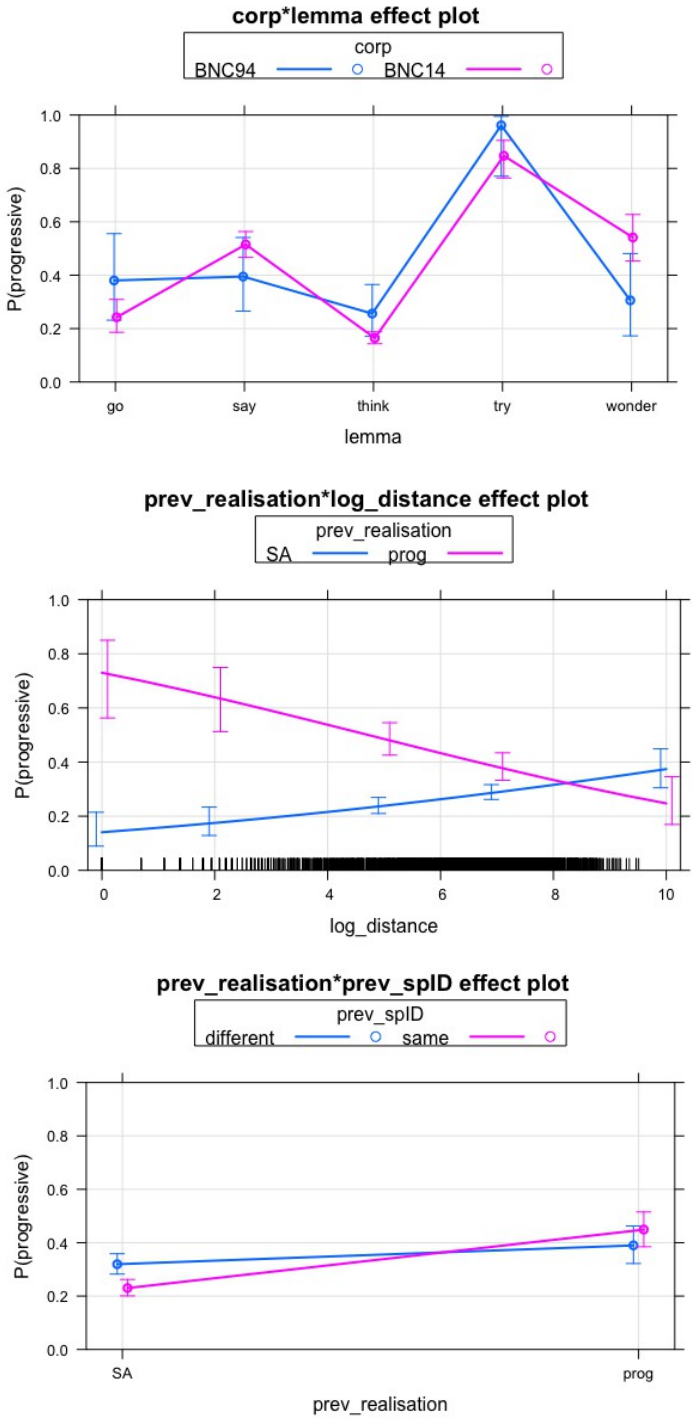

Figure A.14.: Binary logistic regression *I BE just -ing*: Aspectual choice – all effects (SA = simple aspect).

Realisation	freq BNC94	freq BNC94/pmw	freq BNC14	freq BNC14/pmw	% change BNC94/14	sign. level
what I'm saying	128	25.5	280	24.5	–3.9%	*ns*
what you're doing	82	16.4	195	17.1	+4.3%	*ns*
what I'm doing	46	9.2	150	13.1	+42.4%	*
what they're doing	46	9.2	158	13.8	+50%	*
what he's doing	41	8.2	90	7.9	–3.7%	*ns*
what we're doing	39	7.8	93	8.1	+3.8%	*ns*
what you're saying	39	7.8	190	16.6	+112.8%	***
what she's doing	30	6.0	57	5.0	–16.7%	*ns*
what I was thinking	24	4.8	110	9.6	+100%	**
what they're saying	21	4.2	51	4.5	+7.1%	*ns*
what I'm trying	18	3.6	51	4.5	+25%	*ns*
what I was saying	18	3.6	86	7.5	+108.3%	**
what you're talking	16	3.2	88	7.7	+140.6%	***
what he's saying	15	3.0	23	2.0	–33.3%	*ns*
what I was doing	14	2.8	68	6.0	+114.3%	**
what he was doing	13	2.6	51	4.5	+73.1%	*ns*
what they were doing	11	2.2	50	4.4	+100%	*
what you were doing	11	2.2	33	2.9	+31.8%	*ns*
what I'm talking	10	2.0	44	3.9	+95%	*ns*
what we were doing	10	2.0	22	1.9	–5%	*ns*

Table A.7.: Frequency development of the *what PP BE -ing* pattern (20 most frequent realisations) (Fisher's exact test: *** p<0.001, ** p<0.01, * p<0.05) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

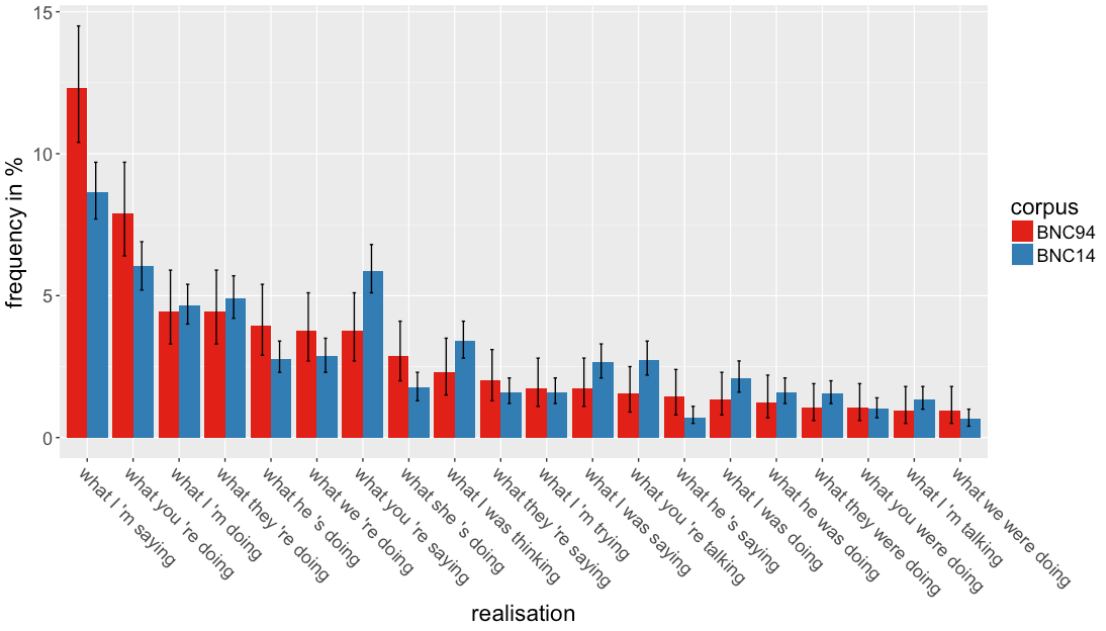

Figure A.15.: Frequency profile of the *what PP BE -ing* pattern (20 most freq. realisations) (freq in %; with 95% confidence intervals) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

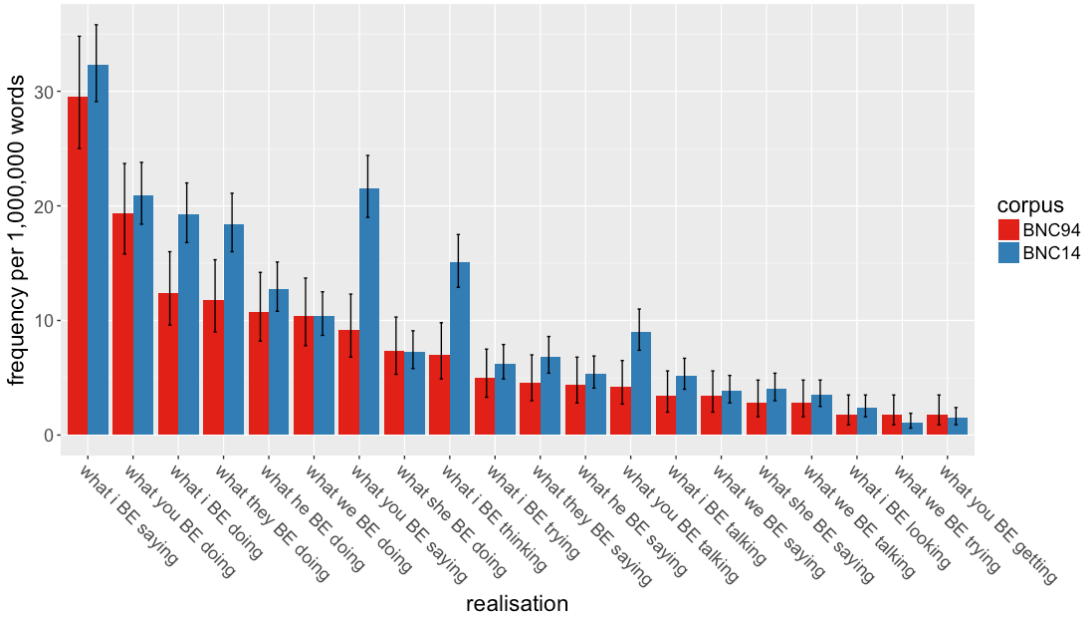

Figure A.16.: Frequency profile of the *what PP BE -ing* pattern (20 most freq. realisations with 95% confidence intervals, no tense distinction) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

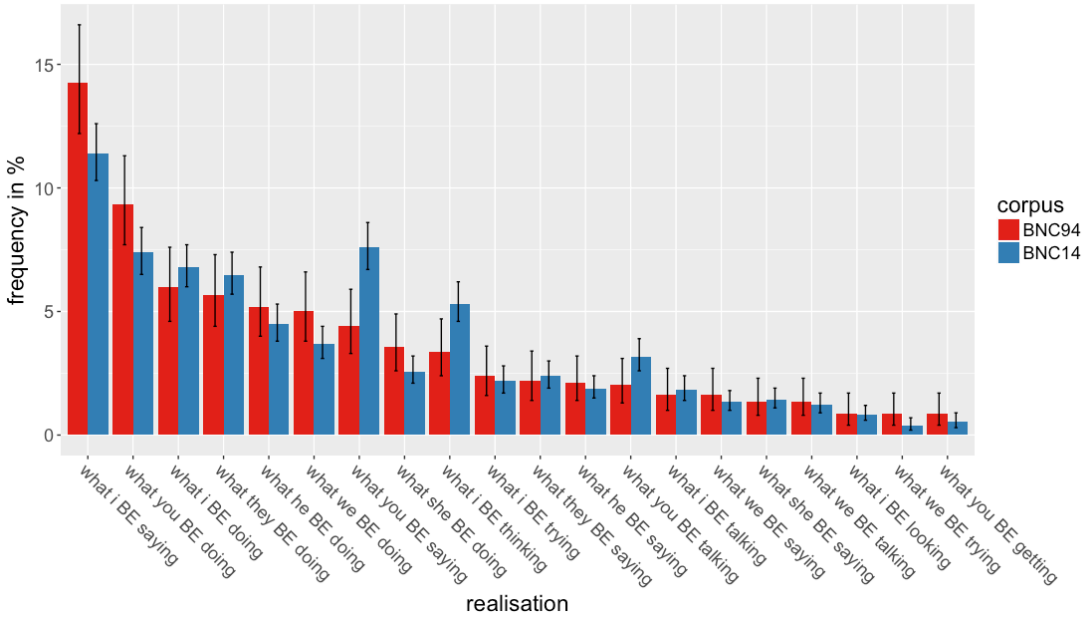

Figure A.17.: Frequency profile of the *what PP BE -ing* pattern (20 most freq. realisations, no tense distinction) (freq in %; with 95% confidence intervals) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

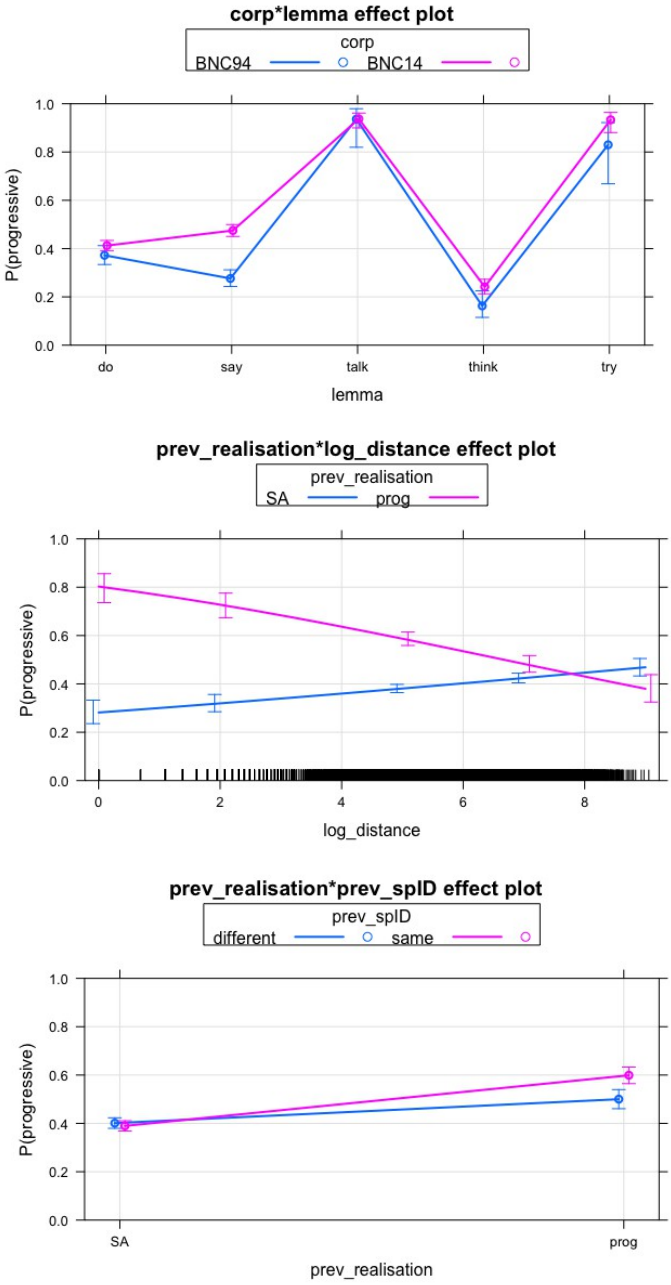

Figure A.18.: Binary logistic regression *what PP BE -ing*: Aspectual choice – all effects (SA = simple aspect).

Realisation	freq BNC94	freq BNC94/pmw	freq BNC14	freq BNC14/pmw	% change BNC94/14	sign. level
what are you doing	227	45.3	496	43.4	–4.2%	*ns*
what are you talking	28	5.6	95	8.3	+48.2%	*ns*
what were you doing	19	3.8	38	3.3	–13.2%	*ns*
what are you saying	16	3.2	33	2.9	–9.4%	*ns*
what are you looking	14	2.8	55	4.8	+71.4%	*ns*
what are you having	11	2.2	23	2.0	–9.1%	*ns*
what're you doing	10	2.0	1	0.1	–95%	***
what were you saying	10	2.0	51	4.5	+125%	*
what are you trying	9	1.8	15	1.3	–27.8%	*ns*
what are you going	8	1.6	4	0.4	–75%	*
what are you making	8	1.6	13	1.1	–31.2%	*ns*
what are you getting	6	1.2	15	1.3	+8.3%	*ns*
what are you eating	5	1.0	9	0.8	–20%	*ns*
what are you playing	5	1.0	8	0.7	–30%	*ns*
what are you giving	4	0.8	1	0.1	–87.5%	*

Table A.8.: Frequency development of the *what BE you -ing* pattern (15 most frequent realisations) (Fisher's exact test: *** p<0.001, ** p<0.01, * p<0.05) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

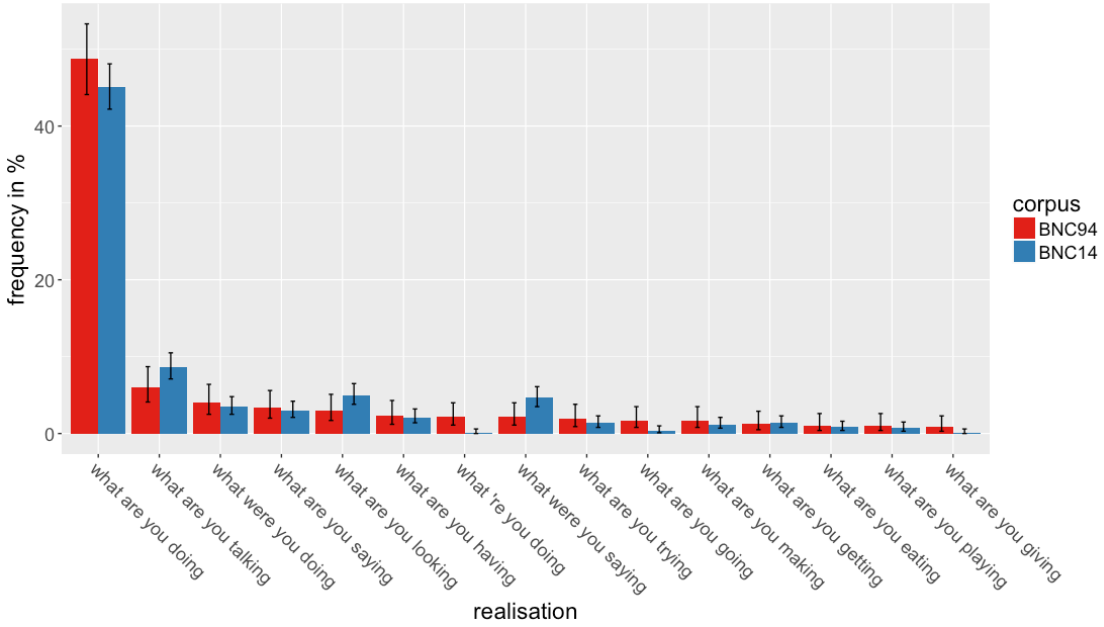

Figure A.19.: Frequency profile of the *what BE you -ing* pattern (15 most freq. realisations) (freq in %; with 95% confidence intervals) (BNC94 = BNC-DS, BNC14 = Spoken BNC2014).

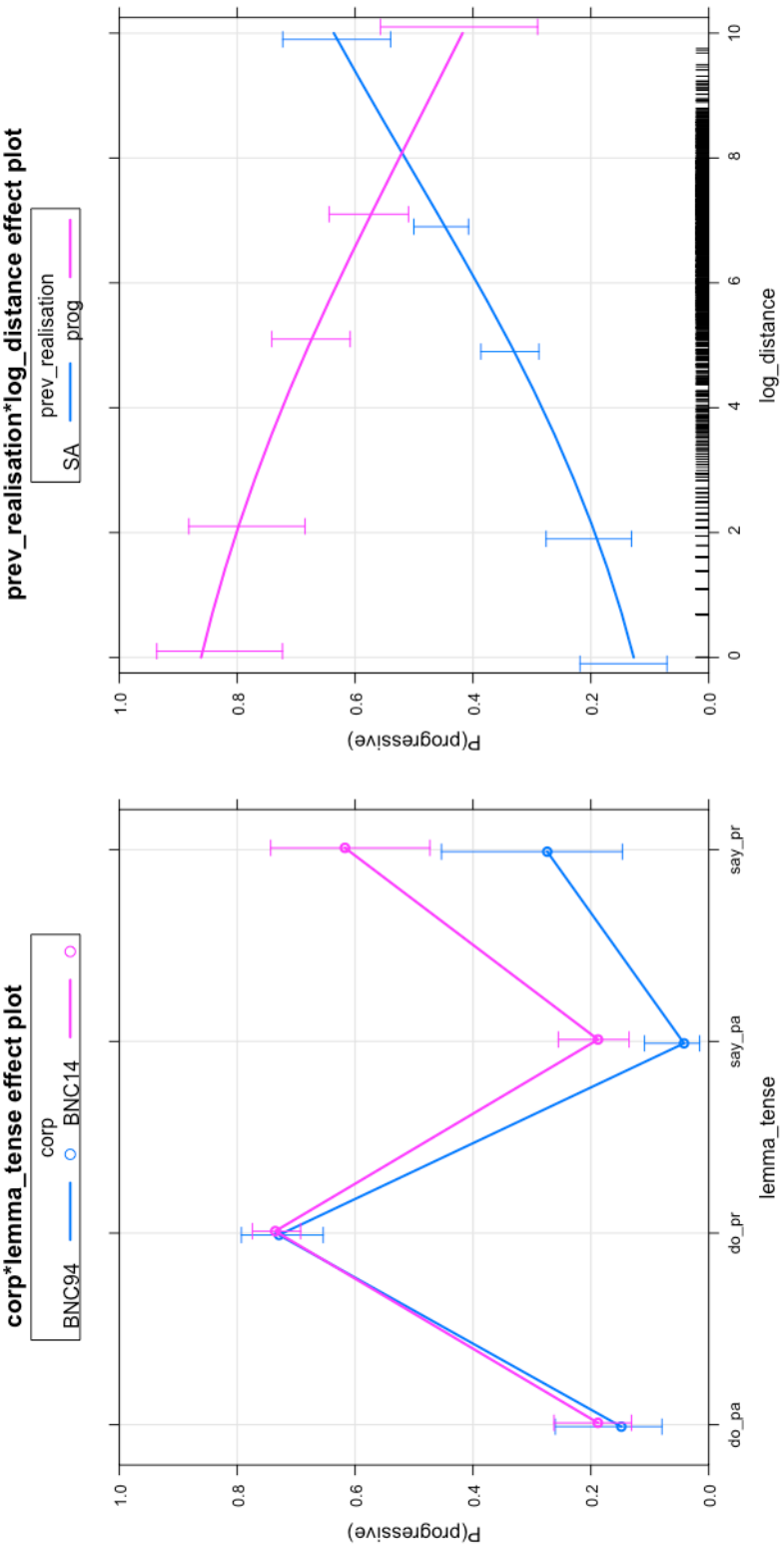

Figure A.20.: Binary logistic regression *what BE you -ing:* Aspectual choice – all effects.

References

Aarts, Bas, Joanne Close, Geoffrey Leech, and Sean Wallis, eds. (2013). *The Verb Phrase in English: Investigating Recent Language Change with Corpora*. Studies in English Language. Cambridge, New York: Cambridge University Press.

Aarts, Bas, Joanne Close, and Sean Wallis (2010). "Recent Changes in the Use of the Progressive Construction in English". In: *Distinctions in English grammar: Offered to Renaat Declerck*. Ed. by Bert Cappelle and Naoaki Wada. Tokyo: Kaitakusha, pp. 148–168.

– (2013). "Choices over Time: Methodological Issues in Investigating Current Change". In: *The Verb Phrase in English: Investigating Recent Language Change with Corpora*. Ed. by Bas Aarts, Joanne Close, Geoffrey Leech, and Sean Wallis. Studies in English Language. Cambridge, New York: Cambridge University Press, pp. 14–45.

Aarts, Bas and Sean Wallis (2006). *The Diachronic Corpus of Present-Day Spoken English*. University College London: Survey of English Usage.

Adolphs, Svenja and Ronald Carter (2013). *Spoken Corpus Linguistics: From Monomodal to Multimodal*. Vol. 15. Routledge Advances in Corpus Linguistics. New York: Routledge, Taylor & Francis Group.

Adolphs, Svenja and Dawn Knight (2010). "Building a Spoken Corpus: What are the Basics?" In: *The Routledge Handbook of Corpus Linguistics*. Ed. by Anne O'Keeffe and Michael McCarthy. Routledge Handbooks in Applied Linguistics. Abingdon, New York: Routledge, pp. 38–52.

Aijmer, Karin (1996). *Conversational Routines in English: Convention and Creativity*. Studies in Language and Linguistics. London, New York: Longman.

Alexander, Marc and Mark Davies (2015). *Hansard Corpus 1803-2005*.

Allison, Paul (2012). *When Can You Safely Ignore Multicollinearity?* Web Log Post. URL: `https://statisticalhorizons.com/multicollinearity` (visited on 03/06/2018).

Ambridge, Ben, Evan Kidd, Caroline F. Rowland, and Anna L. Theakston (2015). "The Ubiquity of Frequency Effects in First Language Acquisition". In: *Journal of Child Language* 42.2, pp. 239–273.

Anderwald, Lieselotte (2017). "I'm Loving It – Marketing Ploy or Language Change in Progress?" In: *Studia Neophilologica* 89.2, pp. 176–196.

Arnaud, René (1983). "On the Progress of the Progressive in the Private Correspondence of Famous British People (1800-1880)". In: *Papers from the Second Scandinavian Symposium on Syntactic Variation*. Ed. by Sven Jacobson. Vol. 57. Stockholm Studies in English. Stockholm: Almqvist&Wiksell, pp. 83–94.

Arnon, Inbal and Neal Snider (2010). "More than Words: Frequency Effects for Multi-word Phrases". In: *Journal of Memory and Language* 62.1, pp. 67–82.

Baayen, R. Harald (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge, New York: Cambridge University Press.

Baker, Paul (2009). "The BE06 Corpus of British English and Recent Language Change". In: *International Journal of Corpus Linguistics* 14.3, pp. 312–337.

Bannard, Colin and Danielle Matthews (2008). "Stored Word Sequences in Language Learning: The Effect of Familiarity on Children's Repetition of Four-Word Combinations". In: *Psychological Science* 19.3, pp. 241–248.

Barðdal, Jóhanna and Spike Gildea (2015). "Diachronic Construction Grammar: Epistemological Context, Basic Assumptions and Historical Implications". In: *Diachronic Construction Grammar*. Ed. by Jóhanna Barðdal, Elena Smirnova, Lotte Sommerer, and Spike Gildea. Vol. 18. Construction Approaches to Language. Amsterdam: John Benjamins, pp. 1–49.

Barðdal, Jóhanna, Elena Smirnova, Lotte Sommerer, and Spike Gildea, eds. (2015). *Diachronic Construction Grammar*. Vol. 18. Construction Approaches to Language. Amsterdam: John Benjamins.

Barlow, Michael and Suzanne Kemmer, eds. (2000). *Usage-based models of language*. Stanford, Calif: CSLI Publications, Center for the Study of Language and Information. 356 pp.

Baxter, Judith and Kieran Wallace (2009). "Outside In-group and Out-group Identities? Constructing Male Solidarity and Female Exclusion in UK Builders' Talk". In: *Discourse & Society* 20.4, pp. 411–429.

Behrens, Heike (2017). "The Role of Analogy in Language Processing and Acquisition". In: *The Changing English Language: Psycholinguistic Perspectives*. Ed. by Marianne Hundt, Sandra Mollin, and Simone E. Pfenninger. Studies in English Language. Cambridge: Cambridge University Press, pp. 215–239.

Behrens, Heike and Stefan Pfänder, eds. (2016). *Experience Counts: Frequency Effects in Language*. Vol. 54. Linguae & Litterae. Publications of the School of Language & Literature. Freiburg Institute of Advanced Studies. Berlin, Boston: De Gruyter.

Bergs, Alexander and Gabriele Diewald (2008). *Constructions and Language Change*. Berlin, New York: Mouton de Gruyter.

Biber, Douglas, Susan Conrad, and Viviana Cortes (2004). "If you look at ...: Lexical Bundles in University Teaching and Textbooks". In: *Applied Linguistics* 25.3, pp. 371–405.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Biber, Douglas and Randi Reppen, eds. (2015). *The Cambridge Handbook of English Corpus Linguistics*. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press.

Blumenthal-Dramé, Alice (2012). *Entrenchment in Usage-based Theories: What Corpus Data Do and Do Not Reveal about the Mind*. Vol. 83. Topics in English Linguistics. Berlin: Mouton de Gruyter.

– (2017). "Entrenchment From a Psycholinguistic and Neurolinguistic Perspective". In: *Entrenchment and the Psychology of Language Learning: How We Reorganize and*

Adapt Linguistic Knowledge. Ed. by Hans-Jörg Schmid. Language and the Human Lifespan. Washington, DC: American Psychological Association, pp. 129–152.

Boas, Hans C. (2013). "Cognitive Construction Grammar". In: *The Oxford Handbook of Construction Grammar*. Ed. by Thomas Hoffmann and Graeme Trousdale. Oxford, New York: Oxford University Press, pp. 233–252.

Bock, Kathryn J. (1986). "Syntactic Persistence in Language Production". In: *Cognitive Psychology* 18.3, pp. 355–387.

Bock, Kathryn J., Gary S. Dell, Franklin Chang, and Kristine H. Onishi (2007). "Persistent Structural Priming from Language Comprehension to Language Production". In: *Cognition* 104.3, pp. 437–458.

Branigan, Holly P., Martin J. Pickering, Simon P. Liversedge, Andrew J. Stewart, and Thomas P. Urbach (1995). "Syntactic Priming: Investigating the Mental Representation of Language". In: *Journal of Psycholinguistic Research* 24.6, pp. 489–506.

Branigan, Holly P., Martin J. Pickering, Andrew J. Stewart, and Janet F. Mclean (2000). "Syntactic Priming in Spoken Production: Linguistic and Temporal Interference". In: *Memory & Cognition* 28.8, pp. 1297–1302.

Brinton, Laurel J. (1988). *The Development of English Aspectual Systems: Aspectualizers and Post-Verbal Particles*. Vol. 49. Cambridge Studies in Linguistics. Cambridge, New York: Cambridge University Press.

Burnard, Lou, ed. (2000). *Reference Guide for the British National Corpus (World Edition)*. Oxford: The BNC Archive.

Bybee, Joan (2003). "Mechanisms of Change in Grammaticization: The Role of Frequency". In: *The Handbook of Historical Linguistics*. Ed. by Brian D. Joseph and Richard D. Janda. Oxford: Blackwell Publishing, pp. 602–623.

– (2006). "From Usage to Grammar: The Mind's Response to Repetition". In: *Language* 82.4, pp. 711–733.

– (2007). *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.

– (2010). *Language, Usage and Cognition*. Cambridge, New York: Cambridge University Press.

– (2013). "Usage-based Theory and Exemplar Representation of Constructions". In: *The Oxford Handbook of Construction Grammar*. Ed. by Thomas Hoffmann and Graeme Trousdale. Oxford, New York: Oxford University Press, pp. 49–69.

– (2015). *Language Change*. Cambridge Textbooks in Linguistics. New York: Cambridge University Press.

Bybee, Joan and Paul J. Hopper, eds. (2001). *Frequency and the Emergence of Linguistic Structure*. Vol. 45. Typological Studies in Language. Amsterdam: John Benjamins.

Bybee, Joan and Carol Lynn Moder (2017). "Chunking and Changes in Compositionality in Context". In: *The Changing English Language: Psycholinguistic Perspectives*. Ed. by Marianne Hundt, Sandra Mollin, and Simone E. Pfenninger. Cambridge: Cambridge University Press, pp. 148–170.

Bybee, Joan, Revere Perkins, and William Pagliuca (1994). *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Bybee, Joan and Joanne Scheibman (1999). "The Effect of Usage on Degrees of Constituency: The Reduction of don't in English". In: *Linguistics* 37.4, pp. 575–596.

Bybee, Joan and Rena Torres Cacoullos (2009). "The Role of Prefabs in Grammaticization: How the Particular and the General Interact in Language Change". In: *Formulaic Language: Volume I. Distribution and Historical Change*. Ed. by Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali, and Kathleen Wheatley. Vol. 82. Typological Studies in Language. Amsterdam: John Benjamins, pp. 187–218.

Caldwell-Harris, Catherine, Jonathan Berant, and Shimon Edelman (2012). "Measuring Mental Entrenchment of Phrases with Perceptual Identification, Familiarity Ratings, and Corpus Frequency Statistics". In: *Frequency Effects in Language Representation*. Ed. by Dagmar Divjak and Stefan Th. Gries. Berlin, Boston: De Gruyter, pp. 165–194.

Calude, Andreea S. (2009). "Formulaic Tendencies of Demonstrative Clefts in Spoken English". In: *Formulaic Language: Volume 1. Distribution and Historical Change*. Ed. by Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali, and Kathleen Wheatley. Vol. 82. Typological Studies in Language. Amsterdam: John Benjamins, pp. 55–76.

Celle, Agnès and Nicholas Smith (2010). "Beyond Aspect: will be -ing and shall be -ing". In: *English Language and Linguistics* 14.2, pp. 239–269.

Christiansen, Morten H. and Inbal Arnon (2017). "More Than Words: The Role of Multiword Sequences in Language Learning and Use". In: *Topics in Cognitive Science* 9.3, pp. 542–551.

Coates, Jennifer (1983). *The Semantics of the Modal Auxiliaries*. Croom Helm Linguistics Series. London: Croom Helm.

Comrie, Bernard (1976). *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

– (1995). "Tense and Aspect". In: *Syntax: An International Handbook of Contemporary Research*. Ed. by Joachim Jacobs and Arnim von Stechow. Berlin, Boston: De Gruyter, pp. 1244–1251.

Corrigan, Roberta, Edith A. Moravcsik, Hamid Ouali, and Kathleen Wheatley, eds. (2009a). *Formulaic Language: Volume 1. Distribution and Historical Change*. Vol. 82. Typological Studies in Language. Amsterdam: John Benjamins.

– eds. (2009b). *Formulaic Language: Volume 2. Acquisition, Loss, Psychological Reality, and Functional Explanations*. Vol. 83. Typological Studies in Language. Amsterdam: John Benjamins.

Croft, William (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford, New York: Oxford University Press.

– (2012). *Verbs: Aspect and Causal Structure*. Oxford Linguistics. Oxford, New York: Oxford University Press.

Croft, William and D. A. Cruse (2004). *Cognitive Linguistics*. Cambridge Textbooks in Linguistics. Cambridge, New York: Cambridge University Press.

Crowdy, Steve (1995). "The BNC Spoken Corpus". In: *Spoken English on the Computer: Transcription, Mark-Up and Annotation*. Ed. by Geoffrey Leech, Greg Myers, and Jenny Thomas. Harlow: Longman, pp. 224–234.

Culicover, Peter W., Ray Jackendoff, and Jenny Audring (2017). "Multiword Constructions in the Grammar". In: *Topics in Cognitive Science* 9.3, pp. 552–568.

Dąbrowska, Ewa (2016). "Cognitive Linguistics' Seven Deadly Sins". In: *Cognitive Linguistics* 27.4, pp. 479–491.

Dahl, Östen (1985). *Tense and Aspect Systems*. Oxford, New York: Basil Blackwell.

Davies, Mark (2009). "The British Component of the International Corpus of English (ICE-GB), Release 2, and: Diachronic Corpus of Present-Day Spoken English (DCPSE), and: The International Corpus of English Corpus Utility Program (ICECUP), Version 3.1 (review)". In: *Language* 85.2, pp. 443–445.

De Groot, Casper (2007). "'The king is on huntunge': On the Relation Between Progressive and Absentive in Old and Early Modern English". In: *Structural-Functional Studies in English Grammar*. Ed. by Mike Hannay and Gerard J. Steen. Vol. 83. Studies in Language Companion Series. Amsterdam: John Benjamins, pp. 175–190.

De Smet, Hendrik (2016). "How Gradual Change Progresses: The Interaction between Convention and Innovation". In: *Language Variation and Change* 28.1, pp. 83–102.

– (2017). "Entrenchment Effects in Language Change". In: *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Ed. by Hans-Jörg Schmid. Language and the Human Lifespan. Washington, DC: American Psychological Association, pp. 75–99.

De Smet, Hendrik and Olga Fischer (2017). "The Role of Analogy in Language Change: Supporting Constructions". In: *The Changing English Language: Psycholinguistic Perspectives*. Ed. by Marianne Hundt, Sandra Mollin, and Simone E. Pfenninger. Studies in English Language. Cambridge: Cambridge University Press, pp. 240–268.

Denison, David (1998). "Syntax". In: *The Cambridge History of the English Language: Volume 4 1776–1997*. Ed. by Suzanne Romaine. Cambridge: Cambridge University Press, pp. 92–329.

Dennis, Leah (1940). "The Progressive Tense: Frequency of its Use in English". In: *PMLA* 55.3, pp. 855–865.

Deshors, Sandra C. (2017). "Zooming in on Verbs in the Progressive: A Collostructional and Correspondence Analysis Approach". In: *Journal of English Linguistics* 45.3, pp. 260–290.

Diessel, Holger (2007). "Frequency Effects in Language Acquisition, Language Use, and Diachronic Change". In: *New Ideas in Psychology* 25.2, pp. 108–127.

– (2011). "Language, Usage and Cognition. By Joan Bybee (review)". In: *Language* 87.4, pp. 830–844.

– (2015). "Usage-based Construction Grammar". In: *Handbook of Cognitive Linguistics*. Ed. by Ewa Dabrowska and Dagmar Divjak. Berlin, München, Boston: De Gruyter, pp. 296–322.

– (2016). "Frequency and Lexical Specificity in Grammar: A Critical Review". In: *Experience Counts: Frequency Effects in Language*. Ed. by Heike Behrens and Stefan Pfänder. Berlin, Boston: De Gruyter, pp. 209–237.

Divjak, Dagmar (2012). "Introduction". In: *Frequency Effects in Language Representation*. Ed. by Dagmar Divjak and Stefan Th. Gries. Berlin, Boston: De Gruyter, pp. 1–10.

Divjak, Dagmar and Catherine Caldwell-Harris (2015). "Frequency and Entrenchment". In: *Handbook of Cognitive Linguistics*. Ed. by Ewa Dabrowska and Dagmar Divjak. Berlin, München, Boston: De Gruyter, pp. 53–75.

Divjak, Dagmar and Stefan Th. Gries, eds. (2012). *Frequency Effects in Language Representation*. Vol. 244.2. Trends in Linguistics: Studies and Monographs. Berlin: Mouton de Gruyter.

Ellis, Nick C. (2012). "What can we Count in Language, and what Counts in Language Acquisition, Cognition, and Use?" In: *Frequency Effects in Language Learning and Processing*. Ed. by Stefan Th. Gries and Dagmar Divjak. Berlin, Boston: De Gruyter, pp. 7–33.

– (2016). "Frequency in Language Learning and Language Change: The Contributions to this Volume from a Cognitive and Psycholinguistic Perspective". In: *Experience Counts: Frequency Effects in Language*. Ed. by Heike Behrens and Stefan Pfänder. Berlin, Boston: De Gruyter, pp. 239–254.

– (2017). "Chunking in Language Usage, Learning and Change: I Don't Know". In: *The Changing English Language: Psycholinguistic Perspectives*. Ed. by Marianne Hundt, Sandra Mollin, and Simone E. Pfenninger. Studies in English Language. Cambridge: Cambridge University Press, pp. 113–147.

Ellis, Nick C. and Dave C. Ogden (2017). "Thinking About Multiword Constructions: Usage-Based Approaches to Acquisition and Processing". In: *Topics in Cognitive Science* 9.3, pp. 604–620.

Ellis, Nick C., Ute Römer, and Matthew Brook O'Donnell (2016). *Usage-based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar*. Chichester: Wiley.

Elsness, Johan (1994). "On the Progression of the Progressive in Early Modern English." In: *ICAME Journal* 18, pp. 5–25.

Field, Andy, Jeremy Miles, and Zoë Field (2012). *Discovering Statistics Using R*. London, Thousand Oaks: Sage.

Fillmore, Charles J. (1988). "The Mechanisms of Construction Grammar". In: *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*. Ed. by Shelley Axmaker, Annie Jaisser, and Helen Singmaster. Berkeley: Berkeley Linguistics Society, pp. 35–55.

Fillmore, Charles J., Paul Kay, and Mary Catherine O'Connor (1988). "Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone". In: *Language* 64.3, pp. 501–538.

Fischer, Olga (1992). "Syntax". In: *The Cambridge History of the English Language: Volume 2 1066–1476*. Ed. by Norman Blake. Cambridge: Cambridge University Press, pp. 207–408.

Ford, Marilyn, Joan Bresnan, and Ronald M. Kaplan (1982). "A Competence-based Theory of Syntactic Change". In: *Mental Representations and Grammatical Relations*. Ed. by Joan Bresnan. Cambridge, MA: MIT Press, pp. 727–796.

Francis, W. Nelson (1992). "The London-Lund Corpus of Spoken English: Description and Research Edited by Jan Svartvik". In: *Language* 68.1, pp. 196–199.

Fried, Mirjam (2013). "Principles of Constructional Change". In: *The Oxford Handbook of Construction Grammar*. Ed. by Thomas Hoffmann and Graeme Trousdale. Oxford, New York: Oxford University Press, pp. 419–437.

Friendly, Michael and David Meyer (2016). *Discrete Data Analysis with R: Visualization and Modeling Techniques for Categorical and Count Data*. Chapman & Hall/CRC texts in Statistical Science Series. Boca Raton: CRC Press, Taylor & Francis Group.

Frost, Jim (2017). *Multicollinearity in Regression Analysis: Problems, Detection, and Solutions*. Statistics By Jim. URL: `http://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/` (visited on 03/06/2018).

Garrod, Simon and Martin J. Pickering (2004). "Why is Conversation so Easy?" In: *Trends in Cognitive Sciences* 8.1, pp. 8–11.

Gilquin, Gaetanelle (2006). "The Verb Slot in Causative Constructions: Finding the Best Fit". In: *Constructions* 1.3, pp. 1–46.

Goldberg, Adele E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Cognitive Theory of Language and Culture. Chicago: University of Chicago Press.

– (2003). "Constructions: A New Theoretical Approach to Language". In: *Trends in Cognitive Sciences* 7.5, pp. 219–224.

– (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford linguistics. Oxford, New York: Oxford University Press.

– (2013). "Constructionist Approaches". In: *The Oxford Handbook of Construction Grammar*. Ed. by Thomas Hoffmann and Graeme Trousdale. Oxford, New York: Oxford University Press, pp. 15–31.

Goldberg, Adele E., Devin M. Casenhiser, and Nitya Sethuraman (2004). "Learning Argument Structure Generalizations". In: *Cognitive Linguistics* 15.3, pp. 289–316.

Gries, Stefan Th. (2004). *HCFA 3.2. A program for R.*

– (2005). "Syntactic Priming: A Corpus-based Approach". In: *Journal of Psycholinguistic Research* 34.4, pp. 365–399.

– (2007). "New Perspectives on Old Alternations". In: *Papers from the 39th Regional Meeting of the Chicago Linguistics Society: Vol. II. The Panels*. Ed. by Jonathan E. Cihlar, Amy L. Franklin, and David W. Kaiser. Chicago: Chicago Linguistics Society, pp. 274–292.

– (2009). *Statistics for Linguistics with R: A Practical Introduction*. Trends in Linguistics. Studies and Monographs 208. Berlin, New York: Mouton de Gruyter.

– (2011). "Studying Syntactic Priming in Corpora: Implications of Different Levels of Granularity". In: *Converging Evidence: Methodological and Theoretical Issues for Linguistic Research*. Ed. by Doris Schönefeld. Vol. 33. Human Cognitive Processing. Amsterdam: John Benjamins, pp. 143–164.

– (2013a). "Data in Construction Grammar". In: *The Oxford Handbook of Construction Grammar*. Ed. by Thomas Hoffmann and Graeme Trousdale. Oxford, New York: Oxford University Press, pp. 93–108.

– (2013b). *Statistics for Linguistics with R: A Practical Introduction*. 2nd ed. Berlin: Mouton de Gruyter.

– (2014). *Coll.analysis 3.5. A Script for R to Compute Collostructional Analyses.*

Gries, Stefan Th. (2015a). "More (old and new) Misunderstandings of Collostructional Analysis: On Schmid and Küchenhoff (2013)". In: *Cognitive Linguistics* 26.3, pp. 505–536.
– (2015b). "The most under-used Statistical Method in Corpus Linguistics: Multi-level (and Mixed-effects) Models". In: *Corpora* 10.1, pp. 95–125.
– (2017a). *Quantitative Corpus Linguistics with R: A Practical Introduction*. 2nd ed. New York: Routledge, Taylor & Francis Group.
– (2017b). *Ten Lectures on Quantitative Approaches in Cognitive Linguistics*. Leiden, Boston: Brill.
Gries, Stefan Th. and Dagmar Divjak, eds. (2012). *Frequency Effects in Language Learning and Processing*. Vol. 244.1. Trends in Linguistics: Studies and Monographs. Berlin, Boston: Mouton de Gruyter.
Gries, Stefan Th. and Nick C. Ellis (2015). "Statistical Measures for Usage-Based Linguistics". In: *Language Learning* 65.1, pp. 228–255.
Gries, Stefan Th., Beate Hampe, and Doris Schönefeld (2005). "Converging Evidence: Bringing Together Experimental and Corpus Data on the Association of Verbs and Constructions". In: *Cognitive Linguistics* 16.4, pp. 635–676.
– (2010). "Converging Evidence II: More on the Association of Verbs and Constructions". In: *Empirical and Experimental Methods in Cognitive/Functional Research*. Ed. by Sally Rice and John Newman. Conceptual Structure, Discourse and Language. Stanford, Calif: CSLI Publications / Center for the Study of Language and Information, pp. 59–72.
Gries, Stefan Th. and Anatol Stefanowitsch (2004). "Extending Collostructional Analysis: A Corpus-based Perspective on 'Alternations'". In: *International Journal of Corpus Linguistics* 9.1, pp. 97–129.
Herbst, Thomas, Susen Faulhaber, and Peter Uhrig, eds. (2011). *The Phraseological View of Language: A Tribute to John Sinclair*. Berlin, Boston: De Gruyter.
Herbst, Thomas, Hans-Jörg Schmid, and Susen Faulhaber, eds. (2014). *Constructions Collocations Patterns*. Vol. 282. Trends in Linguistics: Studies and Monographs. Berlin, Boston: De Gruyter.
Hilpert, Martin (2006). "Distinctive Collexeme Analysis and Diachrony". In: *Corpus Linguistics and Linguistic Theory* 2.2, pp. 243–256.
– (2010). "Wallis, S.A., B. Aarts, G. Ozon, and Y. Kavalova. 2006. The Diachronic Corpus of Present-Day Spoken English (DCPSE)". In: *Corpus Linguistics and Linguistic Theory* 6.1, pp. 105–124.
– (2012). "Diachronic Collostructional Analysis: How to Use it and How to Deal with Confounding Factors". In: *Current Methods in Historical Semantics*. Ed. by Kathryn Allan and Justyna A. Robinson. Berlin, Boston: De Gruyter, pp. 133–160.
– (2013a). *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge: Cambridge University Press.
– (2013b). "Corpus-based Approaches to Constructional Change". In: *The Oxford Handbook of Construction Grammar*. Ed. by Thomas Hoffmann and Graeme Trousdale. Oxford, New York: Oxford University Press, pp. 458–475.

– (2014). *Construction Grammar and its Application to English*. Edinburgh Textbooks on the English Language. Edinburgh: Edinburgh University Press.
– (2017). "Frequencies in Diachronic Corpora and Knowledge of Language". In: *The Changing English Language: Psycholinguistic Perspectives*. Ed. by Marianne Hundt, Sandra Mollin, and Simone E. Pfenninger. Studies in English Language. Cambridge: Cambridge University Press, pp. 49–68.
Hilpert, Martin and Holger Diessel (2017). "Entrenchment in Construction Grammar". In: *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Ed. by Hans-Jörg Schmid. Language and the Human Lifespan. Washington, DC: American Psychological Association, pp. 57–74.
Hoffmann, Sebastian and Stefan Evert (2006). *BNCweb (CQP edition)*.
Hoffmann, Thomas and Graeme Trousdale (2011). "Special Issue: Variation, Change and Constructions in English". In: *Cognitive Linguistics* 22.1, pp. 1–209.
– (2013). "Construction Grammar: Introduction". In: *The Oxford Handbook of Construction Grammar*. Ed. by Thomas Hoffmann and Graeme Trousdale. Oxford, New York: Oxford University Press, pp. 1–12.
Hopper, Paul J. and Elizabeth Closs Traugott (2003). *Grammaticalization*. 2nd ed. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
Horch, Stephanie (2017). *Conversion in Asian Englishes: A Usage-based Account of the Emergence of New Local Norms*. New Ideas in Human Interaction. Freiburg: Universitätsbibliothek Freiburg.
Huddleston, Rodney D. and Geoffrey K. Pullum (2002). *The Cambridge Grammar of the English Language*. Cambridge, New York: Cambridge University Press.
Hundt, Marianne (2004a). "Animacy, Agentivity, and the Spread of the Progressive in Modern English". In: *English Language and Linguistics* 8.1, pp. 47–69.
– (2004b). "The Passival and the Progressive Passive: A Case Study of Layering in the English Aspect and Voice Systems". In: *Corpus Approaches to Grammaticalization in English*. Ed. by Hans Lindquist and Christian Mair. Amsterdam, Philadelphia: John Benjamins, pp. 79–120.
Hundt, Marianne and Christian Mair (1999). "Agile and Uptight Genres: The Corpus-based Approach to Language Change in Progress". In: *International Journal of Corpus Linguistics* 4.2, pp. 221–242.
Hundt, Marianne, Sandra Mollin, and Simone E. Pfenninger, eds. (2017). *The Changing English Language: Psycholinguistic Perspectives*. Studies in English Language. Cambridge: Cambridge University Press.
Hunston, Susan and Gill Francis (2000). *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Vol. 4. Studies in Corpus Linguistics. Amsterdam: John Benjamins.
Israel, Michael (1996). "The Way Constructions Grow". In: *Conceptual Structure, Discourse, and Language*. Ed. by Adele E. Goldberg. Stanford, Calif: CSLI Publications, pp. 217–230.
Jäger, Gerhard and Anette Rosenbach (2008). "Priming and Unidirectional Language Change". In: *Theoretical Linguistics* 34.2, pp. 85–113.

Kapatsinski, Vsevolod and Joshua Radicke (2009). "Frequency and the Emergence of Prefabs: Evidence from Monitoring". In: *Formulaic Language: Volume 2. Acquisition, Loss, Psychological Reality, and Functional Explanations*. Ed. by Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali, and Kathleen Wheatley. Vol. 83. Typological Studies in Language. Amsterdam: John Benjamins, pp. 499–520.

Kay, Paul and Charles J. Fillmore (1999). "Grammatical Constructions and Linguistic Generalizations: The What's X Doing Y? Construction". In: *Language* 75.1, pp. 1–33.

Kemmer, Suzanne and Michael Barlow (2000). "Introduction: A Usage-Based Conception of Language". In: *Usage-based Models of Language*. Ed. by Michael Barlow and Suzanne Kemmer. Stanford, Calif: CSLI Publications, Center for the Study of Language and Information, pp. vii–xxviii.

Kiesling, Scott F. (2018). *Investment in Stancetaking: I mean, Just sayin'*. Presentation given at the 5th International Society for the Linguistics of English Conference. London: ISLE.

Killie, Kristin (2008). "From Locative to Durative to Focalized? The English Progressive and 'PROG Imperfective Drift'". In: *English Historical Linguistics 2006: Volume 1. Syntax and Morphology*. Ed. by Maurizio Gotti, Marina Dossena, and Richard Dury. Vol. 295. Amsterdam Studies in the Theory and History of Linguistic Science. Amsterdam, Philadelphia: John Benjamins, pp. 69–88.

Kirk, John M. and Gisle Andersen (2016). "Compilation, Transcription, Markup and Annotation of Spoken Corpora". In: *International Journal of Corpus Linguistics* 21.3, pp. 291–298.

Klein, Wolfgang (1994). *Time in Language*. Germanic Linguistics. London, New York: Routledge.

– (1995). "A Simplest Analysis of the English Tense-Aspect System". In: *Anglistentag 1994 Graz: Proceedings*. Vol. XVI. Proceedings of the Conference of the German Association of University Teachers of English. Tübingen: Max Niemeyer, pp. 139–151.

König, Ekkehard (1980). "On the Context-Dependence of the Progressive in English". In: *Time, Tense, and Quantifiers: Proceeding of the Stuttgart Conference on the Logic of Tense and Quantification*. Ed. by Christian Rohrer. Vol. 83. Linguistische Arbeiten. Tübingen: Max Niemeyer, pp. 269–291.

– (1995a). "HE IS BEING OBSCURE: Non-Verbal Predication and the Progressive". In: *Temporal Reference, Aspect and Actionality. Vol.2: Typological Perspectives*. Ed. by Pier Marco Bertinetto, Valentina Bianchi, Östen Dahl, and Mario Squartini. Turin: Rosenberg & Sellier, pp. 155–167.

– (1995b). "On Analyzing the Tense-Aspect System of English: A State-of-the-Art Report". In: *Anglistentag 1994 Graz: Proceedings*. Proceedings of the Conference of the German Association of University Teachers of English XVI. Tübingen: Max Niemeyer, pp. 153–169.

Kortmann, Bernd (1991). "The Triad 'Tense-Aspect-Aktionsart': Problems and Possible Solutions". In: *Belgian Journal of Linguistics* 6, pp. 9–30.

Kranich, Svenja (2010). *The Progressive in Modern English: A Corpus-based Study of Grammaticalization and Related Changes*. Vol. 72. Language and Computers: Studies in Practical Linguistics. Amsterdam, New York: Rodopi.

– (2013). “Functional Layering and the English Progressive”. In: *Linguistics* 51.1, pp. 1–32.

Lakoff, George (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago: The University of Chicago Press.

Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar. Vol. 1: Theoretical Prerequisites*. Stanford: Stanford University Press.

– (2000). “A Dynamic Usage-Based Model”. In: *Usage-based Models of Language*. Ed. by Michael Barlow and Suzanne Kemmer. Stanford: CSLI Publications, Center for the Study of Language and Information, pp. 1–63.

Leech, Geoffrey (2000). “Grammars of Spoken English: New Outcomes of Corpus-Oriented Research”. In: *Language Learning* 50.4, pp. 675–724.

– (2004). *Meaning and the English Verb*. 3rd ed. Harlow, New York: Pearson/Longman.

– (2015). “Descriptive Grammar”. In: *The Cambridge Handbook of English Corpus Linguistics*. Ed. by Douglas Biber and Randi Reppen. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press, pp. 146–160.

Leech, Geoffrey, Marianne Hundt, Christian Mair, and Nicholas Smith (2009). *Change in Contemporary English: A Grammatical Study*. Studies in English Language. Cambridge, New York: Cambridge University Press.

Leech, Geoffrey and Nicholas Smith (2005). “Extending the Possibilities of Corpus-Based Research on English in the Twentieth Century: A Prequel to LOB and FLOB”. In: *ICAME Journal* 29, pp. 83–98.

Leech, Geoffrey, Nicholas Smith, and Paul Rayson (2012). “English Style on the Move: Variation and Change in Stylistic Norms in the Twentieth Century”. In: *English Corpus Linguistics: Crossing Paths*. Vol. 76. Language and Computers: Studies in Practical Linguistics. Amsterdam, New York: Rodopi, pp. 69–98.

Lehmann, Christian (1995). *Thoughts on Grammaticalization*. Rev. and expanded version, 1st ed. Vol. 1. LINCOM Studies in Theoretical Linguistics. München: LINCOM Europa.

Levin, Magnus (2013). “The Progressive Verb in Modern American English”. In: *The Verb Phrase in English: Investigating Recent Language Change with Corpora*. Ed. by Bas Aarts, Joanne Close, Geoffrey Leech, and Sean Wallis. Studies in English Language. Cambridge, New York: Cambridge University Press, pp. 187–216.

Ljung, Magnus (1980). *Reflections on the English Progressive*. Vol. 46. Gothenburg Studies in English. Göteborg: Acta Universitatis Gothoburgensis.

Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina, and Tony McEnery (2017). “The Spoken BNC2014: Designing and Building a Spoken Corpus of Everyday Conversations”. In: *International Journal of Corpus Linguistics* 22.3, pp. 319–344.

Lüdeling, Anke and Merja Kytö, eds. (2008). *Corpus Linguistics Volume 1*. Vol. 29/1. Handbooks of Linguistics and Communication Science. Berlin, New York: De Gruyter.

– eds. (2009). *Corpus Linguistics Volume 2*. Vol. 29/2. Handbooks of Linguistics and Communication Science. Berlin, New York: De Gruyter.

Madlener, Karin (2016). “Input Optimization Effects of Type and Token Frequency Manipulations in Instructed Second Language Learning”. In: *Experience Counts: Fre-*

quency Effects in Language. Ed. by Heike Behrens and Stefan Pfänder. Berlin, Boston: De Gruyter, pp. 133–173.

Mair, Christian (2004). "Corpus Linguistics and Grammaticalisation Theory: Statistics, Frequencies, and Beyond". In: *Corpus Approaches to Grammaticalization in English*. Ed. by Hans Lindquist and Christian Mair. Vol. 13. Studies in Corpus Linguistics. Amsterdam: John Benjamins Publishing Company, pp. 121–150.

– (2006a). "Tracking Ongoing Grammatical Change and Recent Diversification in Present-day Standard English: The Complementary Role of Small and Large Corpora". In: *The Changing Face of Corpus Linguistics*. Ed. by Antoinette Renouf and Andrew Kehoe. Vol. 55. Language and Computers: Studies in Practical Linguistics. Amsterdam: Rodopi, pp. 355–376.

– (2006b). *Twentieth-Century English: History, Variation, and Standardization*. Studies in English Language. Cambridge, New York: Cambridge University Press.

– (2009). "Corpora and the Study of Recent Change in Language". In: *Corpus Linguistics Volume 2*. Ed. by Anke Lüdeling and Merja Kytö. Vol. 29/2. Handbooks of Linguistics and Communication Science. Berlin, New York: De Gruyter, pp. 1109–1125.

– (2017). "From Priming and Processing to Frequency Effects and Grammaticalization? Contracted Semi-Modals in Present-Day English". In: *The Changing English Language: Psycholinguistic Perspectives*. Ed. by Marianne Hundt, Sandra Mollin, and Simone E. Pfenninger. Studies in English Language. Cambridge: Cambridge University Press, pp. 191–212.

Mair, Christian and Marianne Hundt (1995). "Why is the Progressive Becoming More Frequent in English? A Corpus-based Investigation of Language Change in Progress". In: *Anglistentag 1994 Graz*. Ed. by Wolfgang Riehle and Hugo Keiper. Tübingen: Max Niemeyer, pp. 247–254.

McCarthy, Michael and Anne O'Keeffe (2009). "Corpora and Spoken Language". In: *Corpus Linguistics Volume 2*. Ed. by Anke Lüdeling and Merja Kytö. Vol. 29/2. Handbooks of Linguistics and Communication Science. Berlin, New York: De Gruyter, pp. 1008–1024.

Michaelis, Laura A. (2004). "Type Shifting in Construction Grammar: An Integrated Approach to Aspectual Coercion". In: *Cognitive Linguistics* 15.1, pp. 1–67.

Mindt, Dieter (2000). *An Empirical Grammar of the English Verb System*. Berlin: Cornelsen.

Mollin, Sandra (2007). "The Hansard Hazard: Gauging the Accuracy of British Parliamentary Transcripts". In: *Corpora* 2.2, pp. 187–210.

Mossé, Fernand (1938). *Histoire de la Forme Périphrastique être + participe présent en Germanique*. Vol. 1-2. Collection Linguistique. Paris: Librairie C. Klincksieck.

Mustanoja, Tauno F (1960). *A Middle English Syntax*. Helsinki: Societé Néophilologique.

Nattinger, James R. and Jeanette S. DeCarrico (1992). *Lexical Phrases and Language Teaching*. Oxford; New York: Oxford University Press.

Nehls, Dietrich (1974). *Synchron-Diachrone Untersuchungen zur Expanded Form im Englischen: Eine Struktural-Funktionale Analyse*. Vol. 19. Linguistische Reihe. München: Max Hueber Verlag.

Nelson, Gerald, Sean Wallis, and Bas Aarts (2002). *Exploring Natural Language: Working with the British Component of the International Corpus of English*. Vol. 29. Varieties of English Around the World. Amsterdam, Philadelphia: John Benjamins.

Nesselhauf, Nadja (2007). "The Spread of the Progressive and its 'Future' Use". In: *English Language and Linguistics* 11.1, pp. 191–207.

Nesselhauf, Nadja and Ute Römer (2007). "Lexical-Grammatical Patterns in Spoken English: The Case of the Progressive with Future Time Reference". In: *International Journal of Corpus Linguistics* 12.3, pp. 297–333.

Núñez Pertejo, Paloma (2004). *The Progressive in the History of English: With Special Reference to the Early Modern English Period: A Corpus-based Study*. Vol. 6. LINCOM Studies in English Linguistics. München: LINCOM.

O'Donnell, Matthew Brook (2008). "From Corpus to Query and Back Again: The Survey of English Usage and ICECUP Paradigm". In: *International Journal of Corpus Linguistics* 13.3, pp. 387–401.

O'Keeffe, Anne and Michael McCarthy, eds. (2010). *The Routledge Handbook of Corpus Linguistics*. Routledge Handbooks in Applied Linguistics. Abingdon, New York: Routledge.

Pawley, Andrew and Francis H. Syder (1983). "Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency". In: *Language and Communication*. Ed. by Jack C. Richards and Richard W. Schmidt. New York: Longman, pp. 191–226.

Petré, Peter (2016a). "Grammaticalization by Changing Co-text Frequencies: Or why [BE Ving] Became the 'Progressive'". In: *English Language and Linguistics* 20.1, pp. 31–54.

– (2016b). "Unidirectionality as a Cycle of Convention and Innovation: Micro-changes in the Grammaticalization of [BE going to INF]". In: *Belgian Journal of Linguistics* 30, pp. 115–146.

– (2017). "The Extravagant Progressive: An Experimental Corpus Study on the History of Emphatic [BE Ving]". In: *English Language and Linguistics* 21.2, pp. 227–250.

Pfaff, Meike, Alexander Bergs, and Thomas Hoffmann (2013). "I was just reading this article - On the Expression of Recentness and the English Past Progressive". In: *The Verb Phrase in English: Investigating Recent Language Change with Corpora*. Ed. by Bas Aarts, Joanne Close, Geoffrey Leech, and Sean Wallis. Studies in English language. Cambridge, New York: Cambridge University Press, pp. 217–238.

Pfänder, Stefan and Heike Behrens (2016). "Experience Counts: An Introduction to Frequency Effects in Language". In: *Experience Counts: Frequency Effects in Language*. Ed. by Heike Behrens and Stefan Pfänder. Berlin, Boston: De Gruyter, pp. 1–20.

Pfänder, Stefan, Heike Behrens, Peter Auer, Daniel Jacob, Rolf Kailuweit, Lars Konieczny, Bernd Kortmann, Christian Mair, and Gerhard Strube (2013). "Erfahrung zählt: Frequenzeffekte in der Sprache - Ein Werkstattbericht". In: *Zeitschrift für Literaturwissenschaft und Linguistik* 169.43, pp. 7–32.

Pickering, Martin J. and Simon Garrod (2004). "Toward a Mechanistic Psychology of Dialogue". In: *Behavioral and Brain Sciences* 27.2, pp. 169–226.

– (2017). "Priming and Language Change". In: *The Changing English Language: Psycholinguistic Perspectives*. Ed. by Marianne Hundt, Sandra Mollin, and Simone E.

Pfenninger. Studies in English Language. Cambridge: Cambridge University Press, pp. 173–190.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik (1985). *A Comprehensive Grammar of the English Language*. London, New York: Longman.

R Core, Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria (http://www.R-project.org/).

Rautionaho, Paula (2014). *Variation in the Progressive: A Corpus-based Study into World Englishes*. Tampere: Tampere University Press.

Rautionaho, Paula and Sandra C. Deshors (2018). "Progressive or not Progressive?: Modeling the Constructional Choices of EFL and ESL Writers". In: *International Journal of Learner Corpus Research* 4.2, pp. 225–252.

Redington, Martin, Nick Chater, and Steven Finch (1998). "Distributional Information: A Powerful Cue for Acquiring Syntactic Categories". In: *Cognitive Science* 22.4, pp. 425–469.

Rissanen, Matti (2000). "Syntax". In: *The Cambridge History of the English Language: Volume 3 1476–1776*. Ed. by Roger Lass. Cambridge: Cambridge University Press, pp. 187–331.

Rödel, Michael (2014). "Mehr als die Summe der einzelnen Teile: Konstruktionen und ihre Einheiten aus diachroner und synchroner Perspektive". In: *Grammatik als Netzwerk von Konstruktionen*. Ed. by Alexander Lasch and Alexander Ziem. Vol. 15. Sprache und Wissen. Berlin, Boston: De Gruyter, pp. 207–223.

Römer, Ute (2005). *Progressives, Patterns. Pedagogy: A Corpus-Driven Approach to English Progressive Forms, Functions, Contexts, and Didactics*. Vol. 18. Studies in Corpus Linguistics. Amsterdam, Philadelphia: John Benjamins.

Römer, Ute and Rainer Schulze, eds. (2009). *Exploring the Lexis–Grammar Interface*. Vol. 35. Studies in Corpus Linguistics. Amsterdam: John Benjamins.

Ruhi, Şükriye, ed. (2014). *Best Practices for Spoken Corpora in Linguistic Research*. Newcastle upon Tyne: Cambridge Scholars Publ.

Scheffer, Johannes (1975). *The Progressive in English*. Vol. 15. North-Holland Linguistic Series. Amsterdam: North-Holland Publ. Co.

Schmid, Hans-Jörg (2010). "Does Frequency in Text Instantiate Entrenchment in the Cognitive System?" In: *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. Ed. by Dylan Glynn and Kerstin Fischer. Berlin, New York: Mouton de Gruyter.

– (2014). "Lexico-grammatical Patterns, Pragmatic Associations and Discourse Frequency". In: *Constructions Collocations Patterns*. Ed. by Thomas Herbst, Hans-Jörg Schmid, and Susen Faulhaber. Berlin, München, Boston: De Gruyter, pp. 239–293.

– (2015). "A Blueprint of the Entrenchment-and-Conventionalization Model". In: *Yearbook of the German Cognitive Linguistics Association* 3.1, pp. 3–25.

– (2017a). "A Framework for Understanding Linguistic Entrenchment and Its Psychological Foundations". In: *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Ed. by Hans-Jörg Schmid. Language and the Human Lifespan. Washington, DC: American Psychological Association, pp. 9–35.

– ed. (2017b). *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Language and the Human Lifespan. Washington, DC: American Psychological Association.

Schmid, Hans-Jörg and Helmut Küchenhoff (2013). "Collostructional Analysis and Other Ways of Measuring Lexicogrammatical Attraction: Theoretical Premises, Practical Problems and Cognitive Underpinnings". In: *Cognitive Linguistics* 24.3, pp. 531–577.

Schmitt, Norbert, ed. (2004). *Formulaic Sequences: Acquisition, Processing, and Use*. Vol. 9. Language Learning and Language Teaching. Amsterdam, Philadelphia: John Benjamins.

Schmitt, Norbert, Sarah Grandage, and Svenja Adolphs (2004). "Are Corpus-derived Recurrent Clusters Psycholinguistically Valid?" In: *Formulaic Sequences: Acquisition, Processing, and Use*. Ed. by Norbert Schmitt. Vol. 9. Language Learning and Language Teaching. Amsterdam, Philadelphia: John Benjamins, pp. 127–151.

Schneider, Ulrike (2014). *Frequency, Chunks and Hesitations: A Usage-based Analysis of Chunking in English*. New Ideas in Human Interaction. Freiburg: Universitätsbibliothek Freiburg.

– (2016). "Hesitation Placement as Evidence for Chunking: A Corpus-based Study of Spoken English". In: *Experience Counts: Frequency Effects in Language*. Ed. by Heike Behrens and Stefan Pfänder. Berlin, Boston: De Gruyter, pp. 61–89.

Simpson-Vlach, Rita and Nick C. Ellis (2010). "An Academic Formulas List: New Methods in Phraseology Research". In: *Applied Linguistics* 31.4, pp. 487–512.

Sinclair, John (1991). *Corpus, Concordance, Collocation*. Describing English Language. Oxford: Oxford University Press.

Siyanova-Chanturia, Anna and Ron Martinez (2015). "The Idiom Principle Revisited". In: *Applied Linguistics* 36.5, pp. 549–569.

Smith, K. Aaron (2007). "The Development of the English Progressive". In: *Journal of Germanic Linguistics* 19.3, pp. 205–241.

Smith, Nicholas (2002). "Ever Moving On? The Progressive in Recent British English". In: *New Frontiers of Corpus Research: Papers from the Twenty First International Conference on English Language Research on Computerized Corpora, Sydney 2000*. Ed. by Pam Peters, Peter Collins, and Adam Smith. Vol. 36. Language and Computers. Amsterdam: Rodopi, pp. 317–330.

– (2003). "A Quirky Progressive? A Corpus-based Exploration of the will + be + -ing Construction in Recent and Present Day British English." In: *Proceedings of the Corpus Linguistics 2003 Conference*. Ed. by D. Archer, P. Rayson, A. Wilson, and A. McEnrey. Vol. 16. UCREL Technical Paper. Lancaster University, pp. 714–723.

– (2005). "A Corpus-based Investigation of Recent Change in the Use of the Progressive in British English". PhD thesis. University of Lancaster.

Smith, Nicholas and Geoffrey Leech (2013). "Verb Structures in Twentieth-Century British English". In: *The Verb Phrase in English: Investigating Recent Language Change with Corpora*. Ed. by Bas Aarts, Joanne Close, Geoffrey Leech, and Sean Wallis. Studies in English language. Cambridge; New York: Cambridge University Press, pp. 68–98.

Smith, Nicholas and Paul Rayson (2007). "Recent Change and Variation in the British English Use of the Progressive Passive". In: *ICAME Journal* 31, pp. 129–129.

Smitterberg, Erik (2005). *The Progressive in 19th-Century English: A Process of Integration*. Vol. 54. Language and Computers. Amsterdam: Rodopi.

Soete, Nicholas (2014). *Going to Futures with Progressive Infinitives in Varieties of English: A Corpus-based Study*. Master's Thesis. University of Gent.

Stefanowitsch, Anatol (2013). "Collostructional Analysis". In: *The Oxford Handbook of Construction Grammar*. Ed. by Thomas Hoffmann and Graeme Trousdale. Oxford, New York: Oxford University Press, pp. 290–306.

Stefanowitsch, Anatol and Susanne Flach (2017). "The Corpus-Based Perspective on Entrenchment". In: *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Ed. by Hans-Jörg Schmid. Language and the Human Lifespan. Washington, DC: American Psychological Association, pp. 101–127.

Stefanowitsch, Anatol and Stefan Th. Gries (2003). "Collostructions: Investigating the Interaction of Words and Constructions". In: *International Journal of Corpus Linguistics* 8.2, pp. 209–243.

Strang, Barbara M. H. (1982). "Some Aspects of the History of the be + ing Construction". In: *Language Form and Linguistic Variation: Papers Dedicated to Angus McIntosh*. Ed. by John Anderson. Amsterdam: John Benjamins, pp. 425–474.

Szmrecsanyi, Benedikt (2005). "Language Users as Creatures of Habit: A Corpus-based Analysis of Persistence in Spoken English". In: *Corpus Linguistics and Linguistic Theory* 1.1, pp. 113–150.

– (2006). *Morphosyntactic Persistence in Spoken English: A Corpus Study*. Vol. 177. Trends in linguistics: Studies and monographs. Berlin, New York: Mouton de Gruyter.

Thompson, Paul (2005). "Spoken Language Corpora". In: *Developing Linguistic Corpora: A Guide to Good Practice*. Ed. by Martin Wynne. AHDS Guides to Good Practice. Oxford: Oxbow Books. Available online from http://ota.ox.ac.uk/documents/creating/dlc/ [Accessed 2017-08-01].

Traugott, Elizabeth Closs (1989). "On the Rise of Epistemic Meanings in English: An Example of Subjectification in Semantic Change". In: *Language* 65.1, pp. 31–55.

– (1995). "Subjectification in Grammaticalisation". In: *Subjectivity and Subjectivisation: Linguistic Perspectives*. Ed. by Dieter Stein and Susan Wright. Cambridge: Cambridge Univerisity Press, pp. 31–54.

– (2008). "Testing the Hypothesis that Priming is a Motivation for Change". In: *Theoretical Linguistics* 34.2, pp. 135–142.

– (2015). "Toward a Coherent Account of Grammatical Constructionalization". In: *Diachronic Construction Grammar*. Ed. by Jóhanna Barðdal, Elena Smirnova, Lotte Sommerer, and Spike Gildea. Vol. 18. Constructional Approaches to Language. Amsterdam: John Benjamins, pp. 51–79.

Traugott, Elizabeth Closs and Graeme Trousdale (2013). *Constructionalization and Constructional Changes*. Vol. 6. Oxford Studies in Diachronic and Historical Linguistics. Oxford: Oxford University Press.

Tremblay, Antoine, Bruce Derwing, Gary Libben, and Chris Westbury (2011). "Processing Advantages of Lexical Bundles: Evidence From Self-Paced Reading and Sentence Recall Tasks: Lexical Bundle Processing". In: *Language Learning* 61.2, pp. 569–613.

Vendler, Zeno (1957). "Verbs and Times". In: *The Philosophical Review* 66.2, pp. 143–160.

Von Eye, Alexander (1990). *Introduction to Configural Frequency Analysis: The Search for Types and Antitypes in Cross-classifications*. Cambridge, New York: Cambridge University Press.

– (2002). *Configural Frequency Analysis: Methods, Models, and Applications*. Mahwah, N.J: Lawrence Erlbaum Associates.

Wichmann, Anne (2008). "Speech Corpora and Spoken Corpora". In: *Corpus Linguistics Volume 1*. Ed. by Anke Lüdeling and Merja Kytö. Vol. 29/1. Handbooks of Linguistics and Communication Science. Berlin, New York: De Gruyter, pp. 187–207.

Wood, David (2015). *Fundamentals of Formulaic Language: An Introduction*. London, New York: Bloomsbury Academic.

Wray, Alison (2002). *Formulaic Language and the Lexicon*. Cambridge, New York: Cambridge University Press.

– (2008). *Formulaic Language: Pushing the Boundaries*. Oxford Applied Linguistics. Oxford: Oxford University Press.

– (2012). "What Do We (Think We) Know About Formulaic Language? An Evaluation of the Current State of Play". In: *Annual Review of Applied Linguistics* 32, pp. 231–254.

– (2017). "Formulaic Sequences as a Regulatory Mechanism for Cognitive Perturbations During the Achievement of Social Goals". In: *Topics in Cognitive Science* 9.3, pp. 569–587.

Wright, Susan (1994). "The Mystery of the Modal Progressive". In: *Studies in Early Modern English*. Ed. by Dieter Kastovsky. Vol. 13. Topics in English Linguistics. Berlin, New York: Mouton de Gruyter, pp. 467–485.

– (1995). "Subjectivity and Experiential Syntax". In: *Subjectivity and Subjectivisation: Linguistic Perspectives*. Ed. by Dieter Stein and Susan Wright. Cambridge, New York: Cambridge University Press, pp. 151–172.

Wulff, Stefanie, Nick C. Ellis, Ute Römer, Kathleen Bardovi-Harlig, and Chelsea J. Leblanc (2009). "The Acquisition of Tense-Aspect: Converging Evidence From Corpora and Telicity Ratings". In: *The Modern Language Journal* 93.3, pp. 354–369.

Wulff, Stefanie, Anatol Stefanowitsch, and Stefan Th. Gries (2007). "Brutal Brits and Persuasive Americans: Variety-specifc meaning construction in the *into*-causative". In: *Aspects of Meaning Construction*. Ed. by Günter Radden, Klaus-Michael Köpcke, Thomas Berg, and Peter Siemund. Amsterdam: John Benjamins, pp. 265–281.

Xiao, Richard (2009). "Theory-driven Corpus Research: Using Corpora to Inform Aspect Theory". In: *Corpus Linguistics Volume 2*. Ed. by Anke Lüdeling and Merja Kytö. Vol. 29/2. Handbooks of Linguistics and Communication Science. Berlin, New York: De Gruyter, pp. 987–1008.

Žegarac, Vladimir (1993). "Some Observations on the Pragmatics of the Progressive". In: *Lingua* 90.1, pp. 201–220.

Ziem, Alexander (2014). "Konstruktionsgrammatische Konzepte eines Konstruktikons". In: *Grammatik als Netzwerk von Konstruktionen*. Ed. by Alexander Lasch and Alexander Ziem. Vol. 15. Sprache und Wissen. Berlin, Boston: De Gruyter, pp. 15–34.

German summary

In der vorliegenden Arbeit werden der Gebrauch und die Entwicklung der Progressiv-Konstruktion (*be* + Partizip Präsens) in gesprochenem britischen Englisch des zwanzigsten und einundzwanzigsten Jahrhunderts untersucht.

Basierend auf Korpusdaten werden folgende Fragestellungen erforscht:

1. Setzt sich der Frequenz- (d.h. Häufigkeits-) Anstieg des Progressivs im einundzwanzigsten Jahrhundert fort? Falls ja, in welchen Genres?

2. Welche Entwicklungen können als treibende Kräfte hinter einem potentiellen Frequenzanstieg ausgemacht werden? Konventionelle aspektuelle Verwendungen der Konstruktion oder Verwendungen in unkonventionellen/innovativen Kontexten?

3. Wie haben sich frequente lexikalisch verfestigte Progressiv-Patterns im Laufe der letzten zwanzig Jahre entwickelt? Ist ihr Gebrauch stets vom Erwerb routinisierter pragmatischer Bedeutungen gekennzeichnet?

Nach einer kurzen Einleitung stellt Kapitel 2 den theoretischen Hintergrund dar, vor dem die Forschungsergebnisse diskutiert werden. Sprache wird als ein gebrauchsbasiertes System verstanden, das auf generellen kognitiven Fähigkeiten und Prozessen wie z.B. Entrenchment, Abstraktion oder Chunking beruht (vgl. Bybee 2010). Im Sinne der Exemplartheorie wird angenommen, dass jede Verwendung des Progressivs, also jedes Progressiv Token, einen mentalen Niederschlag findet und somit die mentale Repräsentation der Konstruktion beeinflusst. Die verschiedenen morphosyntaktischen, semantischen und pragmatischen Verwendungen des Progressivs werden im Sinne der kognitiven Konstruktionsgrammatik (vgl. Goldberg 1995 & 2006) als komplexes mentales Netzwerk organisiert, das Teil des sogenannten Konstruktikons ist. Die Untersuchung von Frequenzeffekten wie Attraktion oder pragmatischer Verstärkung (pragmatic strengthening) ist natürlicherweise Teil einer solchen Theorie von Sprache.

Kapitel 3 stellt den Forschungsstand zur historischen Entwicklung und zum Gebrauch des Progressivs vor. Zum einen wird klar, dass die Konstruktion seit ihrer Entstehung und Grammatikalisierung einem mehr oder weniger kontinuierlichen Frequenzanstieg unterliegt, der bis ins zwanzigste Jahrhundert andauert (vgl. Mair und Hundt 1995). Zum anderen arbeitet das Kapitel die komplexe Funktionalität der Konstruktion heraus, welche weit über rein aspektuelle Verwendungen hinausgeht.

Basierend auf diesen Erkenntnissen werden die Forschungsfragen (s.o.) und die zentralen Arbeitshypothesen formuliert. Es wird angenommen, dass sich der Frequenzanstieg des Progressivs im einundzwanzigsten Jahrhundert fortsetzt und sich von der Face-to-Face-Interaktion auf gesprochensprachliche Genres ausbreitet, die traditionell stärker von

schriftlichen Normen beeinflusst werden (vgl. Mairs (2006b) Konzept der Kolloquialisierung). Zudem wird angenommen, dass die konventionellen Kontexte der Konstruktion (Präsens (aktiv) plus Aktivitäts-/Ereignisverben) statistisch den größten Beitrag zum fortgesetzten Frequenzanstieg leisten.

Hinsichtlich lexikalisch verfestigter Progressiv-Patterns wird davon ausgegangen, dass deren frequenter Gebrauch in spezifischen diskursiven Kontexten mit einem zunehmenden Grad der Routinisierung und dem Erwerb stabiler pragmatischer Bedeutungen einhergeht.

Kapitel 4 stellt ein neues, eigens für die vorliegende Arbeit erstelltes Korpus vor (*Freiburg Corpus of Spoken English*, FCSE), welches das existierende *Diachronic Corpus of Present-Day Spoken English* (DCPSE) (Aarts und Wallis 2006) um aktuelle Daten ergänzt. Somit werden neben der zweiten Hälfte des zwanzigsten Jahrhunderts (DCPSE) auch die ersten fünfzehn Jahre des einundzwanzigsten Jahrhunderts empirisch abgedeckt.

Zudem werden das *British National Corpus* (BNC) und dessen jüngst verfügbar gemachte diachrone Erweiterung, das *Spoken BNC2014* (Love et al. 2017), für die Untersuchung lexikalisch verfestigter Progressiv-Patterns genutzt.

Kapitel 5 behandelt die ersten beiden Forschungsfragen und zeigt, dass sich die positive Frequenzentwicklung des Progressivs tatsächlich fortsetzt. In allen untersuchten Genres ist die Konstruktion im heutigen Englisch häufiger als Mitte des zwanzigsten Jahrhunderts. Die Entwicklung im Genre Face-to-Face-Konversation verläuft jedoch nicht gleichmäßig, sondern ist durch einen Höchststand in den ICE-GB- (1990er) Daten des DCPSE gekennzeichnet. Die Analyse der BNC- und Spoken BNC2014-Daten legt jedoch nahe, dass es sich hierbei um einen möglicherweise nicht repräsentativen Ausschlag nach oben handelt und dass die Konstruktion auch im informellen konversationalen Sprachgebrauch weiterhin auf dem Vormarsch ist. Allerdings ist diese Entwicklung weniger stark ausgeprägt als in den anderen untersuchten Genres (Radio- und TV-Interviews, Parlamentsdebatten, Reden und Predigten und Live-Kommentare).

Zudem kann gezeigt werden, dass es im Falle eines Frequenzanstieges tatsächlich immer die prototypischsten Verwendungen des Progressivs (im Präsens mit Aktivitäts-/Ereignisverben) sind, die statistisch den größten Beitrag zum Zuwachs leisten. Der Gebrauch in weniger prototypischen Kontexten (z.B. mit Perzeptions-/Sensationsverben, in *to*-Infinitiv-Konstruktionen oder mit bestimmten Modalverben) trägt ebenfalls zur Häufigkeitsentwicklung bei, jedoch statistisch gesehen auf einem niedrigeren Niveau. Somit wird auch die zweite Arbeitshypothese bestätigt.

Verbklassen, die reine Zustandsverben umfassen, also z.B. Beziehungen oder Standpunkte beschreiben, spielen in der Entwicklung der Konstruktion statistisch gesehen kaum eine Rolle. Folglich erscheint die Annahme einer Entwicklung des Progressivs hin zu einem rein imperfektiven Aspekt (cf. Comrie 1995: 1245) nicht plausibel.

In Kapitel 6 werden vier der häufigsten lexikalisch verfestigten Progressiv-Patterns (z.B. *I BE just -ing*) im Detail untersucht. Es wird gezeigt, dass sich die Tokenfrequenz bestimmter verbaler Realisierungen dieser Patterns (z.B. *I'm just saying* oder *thinking*) im Laufe der letzten zwanzig Jahre rapide entwickelt hat. Zudem demonstrieren weitere quantitative und qualitative Untersuchungen, dass die entsprechenden Patterns zumeist klare pragmatische Bedeutungen entwickelt haben und einen hohen Routinisierungsgrad

aufweisen.

Durch die konsequente Analyse gesprochensprachlicher Daten macht die vorliegende Arbeit den derzeitigen Entwicklungsstand des Progressivs im britischen Englisch deutlich. Die Konstruktion unterliegt einem andauernden Frequenzanstieg, im Zuge dessen ihr semantischer Kern jedoch weitestgehend stabil bleibt. An den Rändern ihres konstruktionalen Netzwerks finden jedoch merkliche Veränderungen statt (z.B. die zunehmende Integration von Verbklassen wie Perzeptions-/ Sensationsverben). Diese können als Attraktionseffekt interpretiert werden: Mit zunehmender Tokenfrequenz werden innovative/weniger prototypische Verwendungen der Konstruktion wahrscheinlicher (vgl. De Smet 2016).

Die Analyse lexikalisch verfestigter Progressiv-Patterns zeigt, dass grammatikalische Konstruktionen nicht nur auf abstraktester Ebene untersucht und verstanden werden sollten, sondern die Analyse (semi-)fixierter Schemata zusätzliche Einsichten verspricht. Im vorliegenden Fall sind dies vor allem pragmatische Bedeutungen, die sich im Zuge wiederholter Verwendung in bestimmten diskursiven Kontexten herausbilden (pragmatic strengthening) und mitunter einen hohen Grad der Routinisierung erreichen (siehe z.B. die Reduktion von *I'm just saying* zum Diskursmarker *just saying*).

This book presents an up-to-date account of the use and most recent developments of the progressive construction in present-day spoken British English. Drawing on corpus data from the *Diachronic Corpus of Present-Day Spoken English*, the new *Freiburg Corpus of Spoken English*, the *British National Corpus* and the recently released *Spoken British National Corpus 2014*, it is revealed that the progressive's long-term frequency increase is still continuing in the 21st century.

Taking into account the factors genre, morphosyntactic context and the construction's development with different verbs and verb classes, this study not only quantifies the progressive's ongoing frequency development but also accounts for the underlying driving forces. It is shown that increasing progressive frequencies are based on two major developments: first, further expansion of prototypical core uses and, second, increasing readiness to use the progressive in contexts in which its use has not been obligatory so far (for example with perception/sensation verbs).

Furthermore, the present study analyses frequent lexical-grammatical patterns of progressive use (e.g. *I BE just -ing*). Focusing on their recent frequency development and specific discourse-pragmatic meanings, it can be shown that different patterns have reached different degrees of routinisation in the speech community. While some even exhibit signs of advanced or completed constructionalization (e.g. the reduced form *just saying*), others retain much closer ties with the progressive's basic aspectual meaning.

Udo Rohe studied English, Biology and Pedagogy at the University of Heidelberg and the University of Leeds, UK. He received a teaching degree (Staatsexamen) for secondary education (Gymnasium) in 2014. From 2015 to 2018 he was a member of the doctoral research training group "DFG GRK: Frequency effects in language" at the University of Freiburg. This book is a revised version of his dissertation.

Die Publikationsreihe NIHIN – New Ideas in Human Interaction – ist 2012 in Zusammenarbeit der Hermann Paul School of Linguistics (HPSL) und der Universitätsbibliothek Freiburg (UB) entstanden. NIHIN bietet eine moderne, frei zugängliche Plattform für Dissertationen und Korpora. Zur NIHIN-Reihe gehören NIHIN studies and NIHIN stories.

ISBN 978-3-928969-77-2

eucor The European Campus

Universität Basel

UNI FREIBURG